DETERMINATION OF FATTY ACID SYNTHESIS INTERMEDIATES IN *ESCHERICHIA COLI* AND *BACILLUS SUBTILIS*

BY

SWAMINATH SRINIVAS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Biochemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Doctoral Committee:

Professor John E. Cronan Jr., Chair
Professor James A. Imlay
Professor Satish K. Nair
Professor Wilfred A. van der Donk

ABSTRACT

Lipids play crucial roles in maintaining cellular structure and energy storage. Structural lipids in the form of phospholipids constitute almost 10% of the dry cell weight in bacteria with their synthesis requiring 32 moles of ATP per mole of lipid. This significant investment ensures that the flux through the fatty acid biosynthesis (FAS) and related metabolic pathways is very precisely coordinated. A key feature of the FAS pathway is the acyl carrier protein (ACP), which is a small acidic protein that tethers acyl intermediates via a high-energy thioester bond and shuttles them between enzymes. In *Escherichia coli*, ACP is one of the most abundant soluble proteins with about 60000 copies per cell. Despite being the subject of extensive biochemical and structural studies for several decades, a reliable snapshot of ACP-bound species in any organism under different conditions is unavailable. Previously used methods are severely limited in their capacity to differentiate fatty acid intermediates, suffer from poor reproducibility, require elaborate instrumentation and cannot be used in an ideal setting for determining intracellular fluxes.

This dissertation describes a sensitive and facile method to identify and quantify the physiological level of acyl-ACP species in *E. coli* and *Bacillus subtilis* under varying conditions of growth. The method primarily relies on an inert strep-tagged ACP which is purified using a two-step purification strategy utilizing volatile buffers of low pH in order to preserve the thioester bond between the acyl group and ACP. Samples obtained via this method were extremely pure and free of unwanted salts. The full length purified strep-tagged ACP were subjected to ESI mass spectrometry.

Short and medium chain saturated fatty acyl-ACPs as well as long chain saturated and unsaturated fatty acyl-ACP intermediates were observed. Apart from validating widely held ideas about acyl-ACP flux in basic metabolism, acetyl-ACP, a novel intermediate, was also observed in abundance along with much lower levels of holo-ACP than previously thought. Current efforts are focusing on developing and validating a new model for the initiation of fatty acid biosynthesis using a combination of FAS mutants and stable-isotope labelling studies.

During our efforts to construct one such mutant, we identified and characterized the function of a FabG temperature-sensitive mutant missing eight residues at its dimerization interface. Surprisingly, this mutant behaved like a point mutant. We identified interactions within the dimerization/tetramerization interface that compensate for each other and showed that the wildtype FabG predominantly exists as a dimer but functions as a tetramer. The mutation also rendered the enzyme extremely sensitive to calcium *in vitro* and conferred resistance to a calmodulin inhibitor *in vivo*.

In *B. subtilis,* we have determined that the cellular levels of ACP are ten times lower than those in *E. coli.* when normalized to the total protein content per cell. Even under our stringent purification conditions, ACP co-purified with several other proteins, most noticeably FabF. We are currently trying to identify if Bacillus ACP is part of a loose complex. Nevertheless, this method has also successfully resolved all expected fatty acid intermediates associated with ACP including the presence of acetyl-ACP as observed in *E. coli*.

In addition, I also discuss the development of a new set of vectors with IPTG-dependent origins of replication and a robust CRISPR-Cas9 toolkit for *E. coli* based on this vector set.

ACKNOWLEDGEMENTS

असतो मा सद्गमय । तमसो मा ज्योतिर्गमय ।

From ignorance, lead me to the truth. From darkness, lead me to the light

I would like to start by thanking John Cronan, my advisor and my role model for his guidance and immense help over these many years. He has been patient and supportive throughout the course of my research and has given me the complete freedom to pursue my own scientific curiosities. His infectious passion and dedication to science has helped me define myself as a scientist.

I would like to thank all the members of the Cronan lab, both past and present, for being wonderful colleagues, collaborators and friends. I would especially like to thank Quin Christensen, Alexander Smith, Youjun Feng and Steven Lin for getting me started in lab and for answering innumerable questions patiently. I thank my committee members, Dr. James Imlay, Dr. Satish Nair and Dr. Wilfred van der Donk for their helpful discussion of my work. I also thank Dr. Peter Yau and Dr. Brain Imai from the proteomics facility for helping me understand mass spectrometry and for making this project viable.

Finally, I would like to extend a special thanks to my family, especially my parents and my wife for their constant encouragement, understanding and unwavering support from beginning to end. Without you, none of this would have been possible.

TABLE OF CONTENTS

Chapter One

INTRODUCTION

Functional biological membranes are essential for living organisms. Membranes primarily serve a structural role in the form of selectively permeable barriers that control the movement of biomolecules. In bacteria and eukaryotes, the cell membrane is made up of phospholipids, the hydrophobic components of which are composed of fatty acids, while archaea have isoprenoid-derived lipids. In addition to the cell membrane, Gram-negative bacteria contain an outer membrane, the outer leaflet of which is lipopolysaccharide which contain fatty acids as well. These lipid-rich membranes can also be directly involved in a number of cellular processes such as energy generation and storage, cell signaling, adhesion and pathogenesis. As a consequence, the homeostatic maintenance of the cellular membranes and growth require a constant source of fatty acids, the majority of which are synthesized *de novo*. Two predominant fatty acid biosynthetic systems (FAS) exist in nature: The type I or associated system, found in mammals[1,2] in fungi[3], which utilizes a large multifunctional enzyme complex (270 kDa to 2.6 MDa) containing all the activities necessary to synthesize new fatty acids on a single polypeptide chain, and the type II or dissociated system, in bacteria and plants, which has discrete enzymes performing the individual biosynthetic reactions and a small carrier protein (acyl carrier protein or ACP) that shuttles the intermediates between the enzymes. Given this fundamental difference, the enzymes of bacterial FAS are of tremendous pharmaceutical interest. The model organism for studying type II FAS is *Escherichia coli* where the pathway has been characterized in significant detail. Despite this, several questions remain regarding the structural and mechanistic aspects of each of the individual enzymes and how they are regulated under different conditions.

1.1 TYPE II FATTY ACID BIOSYNTHETIC REACTIONS

Type II fatty acid synthesis begins with the carboxylation of acetyl-CoA using free bicarbonate to generate malonyl-CoA. This is a ATP- and biotin-dependent reaction catalyzed by acetyl-CoA carboxylase and is the first committed step of fatty acid synthesis[4]. The malonyl group is transferred from CoA to the thiol of ACP by a malonyl-CoA-ACP transacylase[5] (FabD).

During the initial round of FAS, the malonyl-ACP is condensed with acetyl-CoA to form acetoacetyl-ACP in a Claisen condensation reaction catalyzed by a β-ketoacyl-ACP synthase[6] (KASIII; FabH). The β-keto group of acetoacetyl-ACP is then reduced to a hydroxyl group by the activity of a β-ketoacyl-ACP reductase FabG in a NADPH dependent manner[7,8]. The β-hydroxybutyryl-ACP is then dehydrated by a β-hydroxyacyl-ACP dehydratase FabZ[9] to form *trans*-2-butenoyl-ACP which is reduced to butyryl-ACP by a NADH dependent enoyl-ACP reductase FabI. While the previous reduction and dehydration reactions are reversible, the equilibrium of the FabI reaction lies strongly towards product formation and pulls the cycle forward. At the end of each cycle of FAS, the saturated fatty acyl-ACP is two carbons longer than when it entered the cycle and can renter the cycle via condensation to malonyl-CoA catalyzed by FabB (KASI) or FabF (KASII). This cycle continues till the desired chain length is reached[10]. FabH is predominant at shorter chain lengths (<6-8 carbons) while FabB/FabF catalyze condensations with longer acyl-ACPs.

Though FabB and FabF have significant overlapping β-ketoacyl-ACP synthase activity, only FabF can elongate palmitoyl-ACP to *cis*-vaccenoyl-ACP, necessary for regulation of membrane fluidity

at varying temperatures. However, this activity and hence, FabF, are non-essential. Though more than 90% of FAS is initiated by FabH using acetyl-CoA, FabH itself is not strictly essential. This is possibly due to the remaining 10% of initiation proceeding using acetyl-ACP instead of acetyl-CoA. The acetyl-ACP itself can be synthesized to a small extent by the transacylation of acetyl-CoA by FabH. However, further studies are necessary to elucidate the proteins responsible for producing this acetyl-ACP and then condensing it with malonyl-ACP.

FabB, however, is essential in *E. coli* due to its indispensable role in unsaturated fatty acid synthesis (UFA). The branch point for UFA synthesis occurs when another dehydratase, FabA, specific for β-hydroxydecanoyl-ACP converts it to *trans*-2-decenoyl-ACP. FabA can also catalyze the isomerization of *trans*-2-decenoyl-ACP to *cis*-3-decenoyl-ACP. This *cis* double bond cannot be reduced by FabI and the *cis*-3-decenoyl-ACP synthesized is further elongated by exclusively FabB while retaining the double bond. Other organisms like *Helicobacter pylori* are shown to utilize a unique and distinct backtracking mechanism in UFA synthesis involving the dehydrogenation of decanoyl-ACP to *trans*-2-decenoyl-ACP and its subsequent isomerization to *cis*-3-decenoyl-ACP using the enzyme FabX[11].

Elongation of acyl-ACPs concludes when the desired acyl chain length is achieved. The fatty acid predominant in *E. coli* phospholipids are palmitic acid (C16:0), *cis*-vaccenic acid (C18:1), palmitoleic acid (C16:1) and to a smaller extent myristic acid (C14:0) and myristoleic acid (C14:1)[12]. The final length is, however, determined as a competition between condensation by β-ketoacyl-ACP synthases and phospholipid synthesis. The acyltransferase PlsB transfers acyl groups from ACP to glycerol-3-phosphate to form lysophosphatidic acid and terminates fatty acid

synthesis. A second acyl chain is transferred by PlsC to form the primary precursor for glycerophospholipids (phosphotidylglycerol, phosphotidylethanolamine and cardiolipin in *E. coli*).

In addition to phospholipid synthesis, shorter acyl-ACP intermediates are also siphoned off by enzymes from other synthetic processes. Malonyl-ACP is utilized by BioC, a malonyl-ACP methyltransferase during biotin synthesis to produce methyl malonyl-ACP. This masks the free carboxyl group and allows it to be extended as an odd carbon-length acyl-ACP intermediate by FAS till it reaches the necessary length (seven carbons) for subsequent biotin synthesis enzymes. The octonoate moiety of octonoyl-ACP is transferred by LipB during lipoic acid synthesis[13] to the lipoyl domain of 2-oxoacid dehydrogenases. Acyl-ACPs that are 14-16 carbon long are utilized as substrates to acylate the protoxin form of hemolysin, via the action of the acyltransferase HlyC. Acyl groups from β-hydroxymyristoyl-ACP via LpxC, myristoyl-ACP via LpxM and lauroyl-ACP via LpxL are respectively utilized during Lipid A biosynthesis. Acyl-ACPs of different lengths are also used by acyl-homoserine lactone synthases to produce quorum sensing molecules.

In *E. coli*, acyl-ACPs are intermediates dedicated to the biosynthetic pathway while acyl-CoAs (except acetyl-CoA) are part of the catabolic β-oxidation pathway. Hence, certain enzymes that specifically require acyl-ACPs like those in lipid A synthesis cannot be supplemented with exogenous fatty acids while those in phospholipid synthesis like PlsB/C which can utilize either acyl-ACPs or acyl-CoAs can. This makes fatty acid biosynthesis absolutely essential in an organism like *E. coli.*

1.2 THE ACYL CARRIER PROTEIN

The acyl carrier protein is a common feature of both type I and type II fatty acid biosynthesis. In the type I system, it exists as a domain of the large polypeptide complex, while in the type II system it is found as a discrete protein that is covalently bound to all acyl intermediates until they are finally transferred to their final recipients. Analogous proteins called PKS ACPs and PCPs are also found in polyketide synthesis and non-ribosomal polypeptide synthesis, respectively, and perform a very similar function[14].

The ACP of *E. coli* is a small acidic protein with a molecular mass of 8.86 kDa and a low isoelectric point of 4.1. This protein is extremely soluble and can be routinely resolubilized in an active form even after harsh protein precipitation/denaturation treatments. Most type II ACPs are strongly conserved and contain four alpha helices. Helix I is antiparallel to helix II and helix IV, which are parallel to each other and linked via the very short Helix III which is perpendicular to both helices. A specific conserved serine (Ser 36 in *E. coli*) present on helix II is modified post-translationally with a 4'-phosphopantetheine moiety (4'-PP) through a phosphodiester bond. This 4'-PP moiety contains a terminal thiol to which the acyl intermediates are covalently attached as a thioester. In *E. coli* a specific and essential holo-ACP synthase AcpS attaches the 4'-phosphopantetheine group derived from coenzyme A to its apo-ACP substrate[15].

This four-helix bundle is flexible and dynamic in solution. This is seen in high-resolution X-ray crystallographic data of apo, holo and acylated forms of various ACPs along with solution state NMR analyses. These crystal structures have revealed that Helix I, II and IV are amphipathic and

have very hydrophobic inner faces and very acidic outer faces and form a hydrophobic sleeve between helix II and helix IV that can sequester the acyl chain and protect it from the solvent[16]. In fact, simply having the 4'-PP group offers a degree of stability as evidenced by reduction in the hydrodynamic radius of holo-ACP in comparison to apo-ACP. The degree of stabilization of the ACP structure is seen to increase with acyl chains up to eight carbons in length suggesting increased interactions with the hydrophobic pocket and indirectly providing a rough physical estimation of the size of the hydrophobic pocket itself. Accommodation of longer acyl chain length is explained by an expanding hydrophobic pocket that pushes helix I away. This, however, comes at the cost of thioester bond stability because longer chains can increase solvent access to the thioester bond. Helix III itself has been seen in a multitude of conformations, and its proximity to the thioester bond suggests that its movements might be critical to accessing proximal parts of the acyl chain.

Co-crystals of ACP with interacting enzymes like AcpS, FabA, LpxD, BioI and BioC suggest that hydrophilic interactions between specific parts of ACP and an arginine/lysine rich 'positive patch' on the interacting enzyme mediate recognition between the proteins. Salt bridges from the acidic residues of helix II of ACP are thought to be primarily involved in this interaction. However, similar salt bridges have also been observed from helix I and IV in other instances. These structures also suggest that to access the acyl chains very little structural change is required, and a transient destabilization of the amphipathic helices of ACP is sufficient to partition the acyl chains into the active sites of the cognate protein. This 'chain flipping' mechanism is spontaneous, and the location of the acyl chain is most likely determined by the relative protection offered by the

hydrophobic pockets of either ACP or its cognate enzyme. This mechanism would probably also extend to PKS ACPs and PCPs since they retain a similar helical bundle fold.

1.3 INTRACELLULAR TURNOVER OF ACP PROSTHETIC GROUP

ACP itself is one of the most abundant monomeric proteins in *E. coli*. More than 60,000 molecules of ACP exist per cell, constituting more than 0.25% of the total soluble protein. The only other proteins present in this scale understandably belong to the translation machinery.

The 4'-phosphopantetheine group of ACP is known to undergo metabolic turnover independent of the protein. This was first shown by selectively incorporating radiolabeled 4'-phosphopantetheine in the intracellular ACP and CoA pools in pantothenate auxotrophs. In *E. coli*, the major enzyme that catalyzes the attachment of the prosthetic group is AcpS. This is a dimeric 4'-phosphopantetheine transferase that transfers the 4'-phosphopantetheine group from coenzyme A to apo-ACP. *E. coli* AcpS is generally specific to its substrates and does not function with the structurally related polyketide and non-ribosomal polypeptide synthesis. Also, AcpS can utilize acetyl-ACP instead of ACP and butyryl-, acetoacetyl- and malonyl-CoAs instead of CoA. However, its low levels in *E. coli* suggest that this activity might not physiologically replace an enzyme like FabD in fatty acid synthesis.

Another enzyme, AcpH, removes the 4'-phosphopantethiene moiety of ACP. Though AcpH is non-essential in *E. coli*, deletion of this gene leads to metabolic instability of the prosthetic group. The exact physiological role of AcpH remains unknown. Thomas and Cronan showed that the

amount of AcpH present in the cell can turn over the entire ACP pool in less than one minute. 4'-phosphopantetheine is found to be excreted by exponential and stationary phase cells in an irreversible process with most of it derived from ACP turnover and very little from CoA turnover.

1.4 COMPOSITION AND REGULATION OF THE INTRACELLULAR ACP POOL

The composition of the intracellular pool of apo-, holo- and acyl-ACPs and the changes in the relative amounts of each of these species under different scenarios of fatty acid regulation has been the subject of some study. Given that fatty acid synthesis is an essential but energy intensive process with roughly 1 ATP and 1 NADPH required per two carbon extension, multiple levels of redundancy and regulation are expected to ensure that growth and cellular processes can proceed in an uninterrupted manner.

A pool of apo-ACP might be expected either as an intermediate of holo-ACP synthesis or as a consequence of turnover of the 4'-phosphopantethiene group discussed before. However, multiple lines of evidence have failed to demonstrate any significant apo-ACP pool under physiological conditions. Given that the intracellular concentration of CoA is roughly 8-10 times more than that of ACP, this is not surprising. Surprisingly, CoA starvation leads to a reduction in active ACP pools rather than an increase in apo-ACP. This was later demonstrated to be a function of amino acid shortage caused by CoA starvation of the TCA cycle resulting in lower *acpP* transcription. Though overexpression of apo-ACP is toxic in *E. coli*, this has been shown to be due to PlsB inhibition rather than an alteration of phospholipids themselves.

Acyl-ACPs were first thought to represent more than 20% of the total intracellular ACP pool in *E. coli*. This was based on the fact that this pool was resistant to alkylation unless first treated with neutral hydroxylamine. Later work showed that this was closer to 12% in rapidly growing cells. Any significant accumulation of acyl-ACPs in the cell has also shown to inhibit fatty acid synthesis. Several fatty acid enzymes have also demonstrated inhibition by long-chain acyl-ACPs including acetyl-CoA carboxylase, the enoyl reductase FabI and the β-ketoacyl-ACP synthase FabH. This inhibition can also be triggered by blocking phospholipid synthesis and can be relieved by overexpressing thioesterases.

Understanding the composition of the ACP pool could give clues towards rate-limiting steps or regulatory points in membrane lipid biogenesis. This regulation can happen either during the early stages of fatty acid biogenesis, by prosthetic group turnover as discussed before or at the point at which fatty acid biosynthesis feeds into phospholipid biosynthesis. Each of these scenarios suggests a completely different ACP pool. If early stages of fatty acid biogenesis are rate limiting, an increased accumulation of acetyl-ACP or malonyl-ACP or even holo-ACP would be expected. With the later stages, an accumulation of long chain acyl-ACPs are expected.

Since the acyltransferase of phospholipid synthesis does not select from a large heterogenous pool of acyl-ACP species, it is unlikely that it decides the fatty acid composition. Instead, the modulation of fatty acid composition observed in various mutants of fatty acid biosynthesis argues that the regulatory control lies with fatty acid synthetic enzymes.

1.5 THE INTERACTIVE NATURE OF ACP

The acyl carrier protein undoubtedly plays a crucial role in fatty acid biogenesis, and its essentiality and intimate interaction with every member of this pathway has been shown in multiple ways including enzymatically and through structural data. Given the cellular abundance of ACP and the relatively low amount of its acylated form in the cell, it has been frequently speculated that ACP might play a secondary role in the cell.

In addition to its demonstrated role in fatty acid associated and dependent processes like biotin and lipoic acid production, lipid A synthesis and phospholipid synthesis, ACP has also been shown to interact with SpoT and mediate guanosine tetraphosphate (ppGpp) levels [17]. ACP has also been shown to associate with membrane associated proteins like YchM [18] to mediate the acquisition of fatty acid precursors. While a lot is known about chain length specificities in specific pathways, a large number of basic questions regarding bacterial metabolism still remain. For example, even though the octanoyltransferase LipB has been characterized to specifically accept eight carbon length acyl-ACPs, the mechanisms guiding their scavenging activity is just being understood.

Mechanism-based inhibitors like electrostatic cross linkers have been used in the past to treat *E. coli* cellular lysates and capture ACP-containing protein complexes. Similar phosphopantetheine based probes have also been employed *in vivo* by hijacking the CoA biosynthetic machinery. In fact, ACP could be identified using cycloaddition-based click chemistry or iterative mass spectrometric approaches that eject the prosthetic group from ACP before detecting them.

Newer studies aimed at finding novel ACP interacting partners have largely used an affinity-based approach with the Tandem Affinity Purification (TAP) method being most popular. This requires co-purification of ACP and its interacting partner with little to no control over the ACP form used to bait the specific partners. Since a majority of the methods rely on the overexpression of the holo-form of ACP (via the co-expression of *acpP* and *acpS*), it is conceivable that a number of partners that require specific acylation of the ACP are missed. Regardless, in a large scale interactome study undertaken in *E. coli* [19], ACP was found to interact with GlmU (lipid A synthesis), AidB (isovaleryl-CoA dehydrogenase), SecA (pre-protein translocase) and MukB (chromosomal partitioning) in addition to the partners indicated before. The physiological importance of these interactions is unclear.

1.6 FATTY ACID SYNTHESIS INITIATION AND ITS INHIBITORS

Initiation of fatty acid synthesis begins with the condensation of malonyl-ACP with acetyl-CoA by the activity of a β-ketoacyl-ACP synthase. Most of the initiation (90%) in *E. coli* proceeds through the activity of FabH, which utilizes acetyl-CoA. Given that FabH shares a similar structural fold with FabB and FabF, it could be hypothesized that one of the three enzymes could condense acetyl-ACP with malonyl-ACP. While FabB and FabF contain a His-His-Cys active site triad and an extended hydrophobic tunnel, structurally FabH only has a small pocket sufficient to bind acetyl-CoA along with a His-Asn-Cys active site. The active site itself is divided into two parts in all three enzymes: one-part binds malonyl-ACP and the other is used for forming an acyl-enzyme intermediate. Two fungal compounds, cerulenin and thiolactomycin, can bind within the active site in distinct ways.

Cerulenin forms a covalent attachment with the active site cysteine of FabB and FabF in an irreversible manner. FabH is refractory to cerulenin, probably due to the lack of a hydrophobic tunnel of suitable size. Thiolactomycin, on the other hand, reversibly inhibits only type II FAS with all three enzymes being sensitive to this compound. Thiolactomycin acts by competing with malonyl-ACP binding. Overexpression of FabB makes *E. coli* resistant to thiolactomycin, indicating that FabB activity can bypass the other initiation pathways.

The model Gram-positive organism *Bacillus subtilis* contains mostly iso and anteiso branched chain fatty acids of 14-17 carbon length in its phospholipids. They utilize acyl-CoA esters of isovalerate, isobutyrate and 2-methylbutyrate as primers instead of acetyl-CoA. These are derived from valine, leucine and isoleucine amino acid catabolic pathways utilizing a specialized branched-chain-keto-acid dehydrogenase complex. *B. subtilis* has two FabH homologs that can accept branch chain primers and initiate fatty acid synthesis. These can also utilize acetyl-CoA to a smaller extent. Thiolactomycin is shown to have very little effect on the branched chain synthases of *Bacillus subtilis* while cerulenin is shown to inhibit *B. subtilis* FabF.

1.7 BETA-KETOACYL-ACP REDUCTASE: FABG

β-ketoacyl-ACP reductases catalyze the reversible reduction of β-ketoacyl-ACP into β-hydroxyacyl-ACP in a cofactor dependent manner and are members of the short-chain dehydrogenase/reductase (SDR) family. In *E. coli*, this activity is encoded by *fabG*, an essential NADPH dependent protein. *fabG* is present within the fatty acid biosynthetic cluster and is co-

transcribed with *plsX*, *fabH*, *fabD* and *acpP*. FabG shows a preference for acyl-ACPs over acyl-CoAs but is nonspecific for the carbon length of the acyl chain. FabG also functions with a variety of substrates including saturated and unsaturated fatty acid intermediates, methylated intermediates in biotin synthesis and even fluorescent molecules.

FabG is easily overexpressed and purified in *E. coli* with very little toxicity. X-ray crystal structures of FabG from *E. coli* with NADPH have been determined at 2.05 A resolution. *E. coli* FabG contains a typical Rossmann fold structure with a twisted parallel beta sheet composed of seven beta strands flanked on both sides by eight alpha helices. The native structure of FabG is identified to be a dimer of dimers. Each monomer is bound to a molecule of $NADP^+$ and two calcium atoms. It has been shown that NADPH binding is accompanied with conformational changes that organizes the active site residues Ser138, Tyr151 and Lys155, indicating allostery.

The first temperature sensitive (Ts) mutants of *fabG* were isolated and characterized to be near the subunit interfaces of the FabG homotetramer. These mutant proteins were extremely thermolabile and could not be purified to homogeneity. Thus, it was hypothesized but never confirmed that these mutants were deficient in tetramerization in a temperature dependent manner. FabG(Ts) mutants accumulated butyryl-ACP and produced very little long chain acyl-ACP intermediates. Members of the SDR family show a strong conservation of structure despite having very low sequence homology. In fact, the conserved SDR folding pattern allows sequence specific assignment of the cofactor binding motif and the active site. FabG is also closely related to another reductase in fatty acid synthesis, the enoyl-ACP reductase, FabI. FabI is primarily NADH

dependent. The equilibrium of the FabI reaction lies strongly towards reduced product formation and thus pulls the FabG reaction to completion in the context of fatty acid synthesis.

One of the intermediate products of fatty acid synthesis from FabG, β-hydroxymyristoyl-ACP, is utilized in one of the first few steps of lipid A synthesis. The enzyme, LpxC, competes with FabZ for the β-hydroxymyristoyl-ACP produced by FabG, indirectly acting as a point of regulation between fatty acid and lipid A synthesis. In fact, the essentiality of lipid A synthesis has made it the target of several antimicrobial compounds. However, surprisingly only FabZ mutants with lower activity have been identified upon treatment of *E. coli* with these compounds.

FabG itself is also of significant interest as an antimicrobial drug target since FabG homologs have been identified in every bacterial genome sequenced to date. Compounds targeting FabG would have a very broad range. However, no inhibitors of FabG have been reported so far. The biggest limitation is the functional flexibility of the FabG active site which makes it difficult to model inhibitors in a flexible binding pocket.

1.8 AIMS AND SCOPE OF THIS THESIS

Lipids are essential components of all living organisms. Synthesis of these lipids is an energy intensive process that is tightly regulated to meet the current physiological demand. Since lipids make up critical structural components of the cell, lipid biosynthesis is also flexible in the range of products it needs to produce to maintain membrane homeostasis. In bacteria, lipids are fatty acid derivatives and are synthesized by the type II fatty acid synthetic system. A number of

intermediates of this pathway are, however, required to sustain other critical biosynthetic pathways like biotin and lipoic acid synthesis, lipid A synthesis and quorum sensing. This ensures that a very complex system of regulation exists to careful channel metabolic flux in a systematic manner. Though we have made great strides in understanding this regulatory mechanism in specific scenarios, a broad and holistic metabolic model is yet to be developed.

Chapter 2 of this thesis describes a method developed to isolate and identify the intermediates of fatty acid biosynthesis in a facile and unbiased manner from *Escherichia coli* and *Bacillus subtilis*. This method relies on the fact that all fatty acid intermediates are naturally covalently tethered to the acyl carrier protein via a thioester linkage. A snapshot of the acyl-ACP pool of both organisms has been obtained under varying metabolic conditions. Acetyl-ACP is identified to be the predominant species of acyl-ACP in both organisms. As a consequence, the interactome of *B. subtilis* ACP has been characterized and its abundance relative to *E. coli* is shown to be 5-fold lower.

Chapter 3 investigates the critical role of dimerization and tetramerization in *E. coli* FabG through the characterization of a temperature sensitive (Ts) version of FabG. This FabG mutant is unique in that the temperature sensitivity is caused due to an eight amino acid deletion unlike other Ts point mutants. The predominant form of *E. coli* FabG is now correctly attributed to its dimer state with tetramerization required before activity. This mutant was re-created in a clean background and found to have aberrant lipid A synthesis and resistance to the drug trifluoperazine.

Chapter 4 describes a number of novel vector systems and tools developed to facilitate studies in Chapter 1 and 2. This includes a convenient and robust CRISPR-Cas9 toolkit that has also been developed based on the above vector backbone. Unlike existing CRISPR-Cas9 systems, these can be easily constructed, tightly regulated and rapidly cured in a reliable manner. To build the system a set of cloning vectors with IPTG dependent origins of replication from multiple incompatibility groups were constructed. These are shown to be maintained in different copy numbers in the presence of IPTG and rapidly lost when IPTG is withdrawn.

Chapter 5 summarizes the results from the preceding chapters and discusses conclusions made from those results as well as work that needs to be done.

1.9 FIGURES

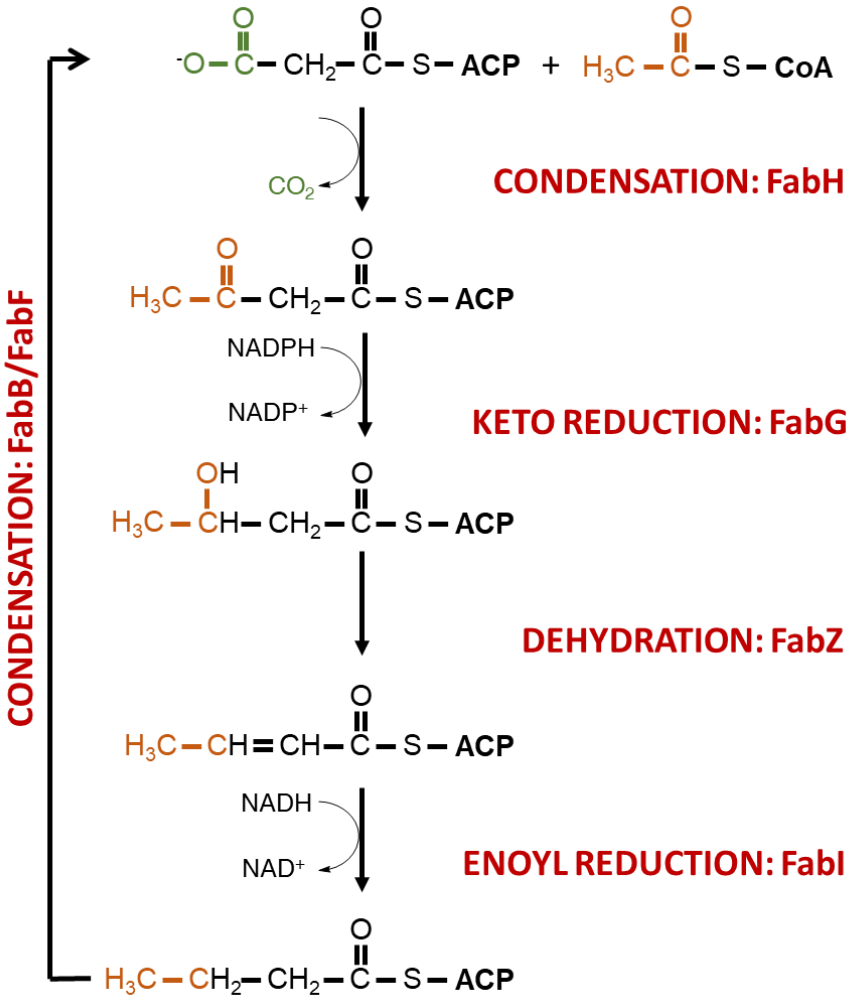

Figure 1.1 The fatty acid biosynthetic reactions. The canonical initiation and elongation steps of fatty acid biosynthesis are represented. The carboxyl group of malonyl-CoA is depicted in green while the acetyl group is in orange to illustrate the head to tail condensation that initiates fatty acid synthesis. The fatty acid synthetic enzymes of *E. coli* are FabB, FabH and FabF: β-ketoacyl-ACP synthase; FabG: β-ketoacyl-ACP reductase; FabZ: β-hydroxyacyl-ACP dehydratase; FabI: enoyl-ACP reductase; ACP: Acyl carrier protein.

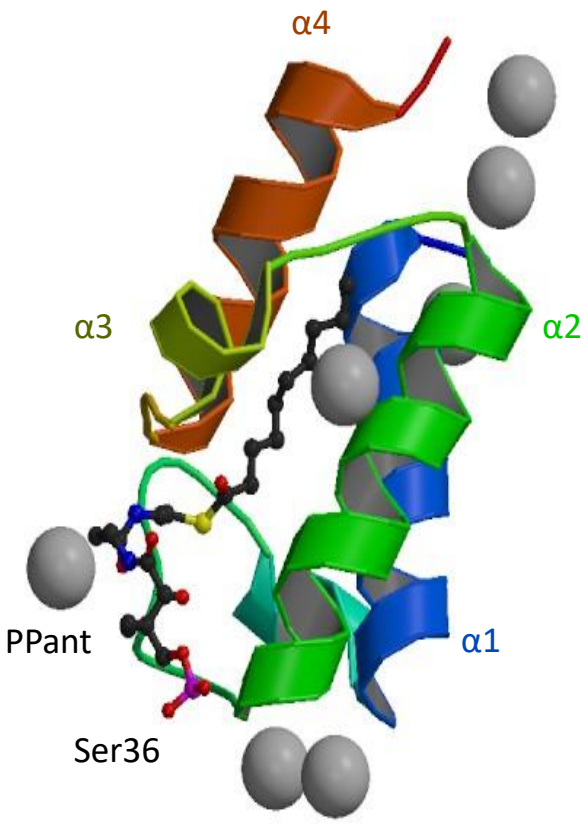

Figure 1.2 Ribbon representation of the structure of decanoyl-ACP. X-ray crystal structure of *E. coli* ACP (PDB 2FAE) with its characteristic four helix bundle. The conserved serine residue (Ser36) is indicated on Helix 2 (in green). This is attached to a 4'-phosphopatetheine via a phosphodiester bond. The decanoyl thioester is sequestered with the hydrophobic cavity formed by helix I, II and IV.

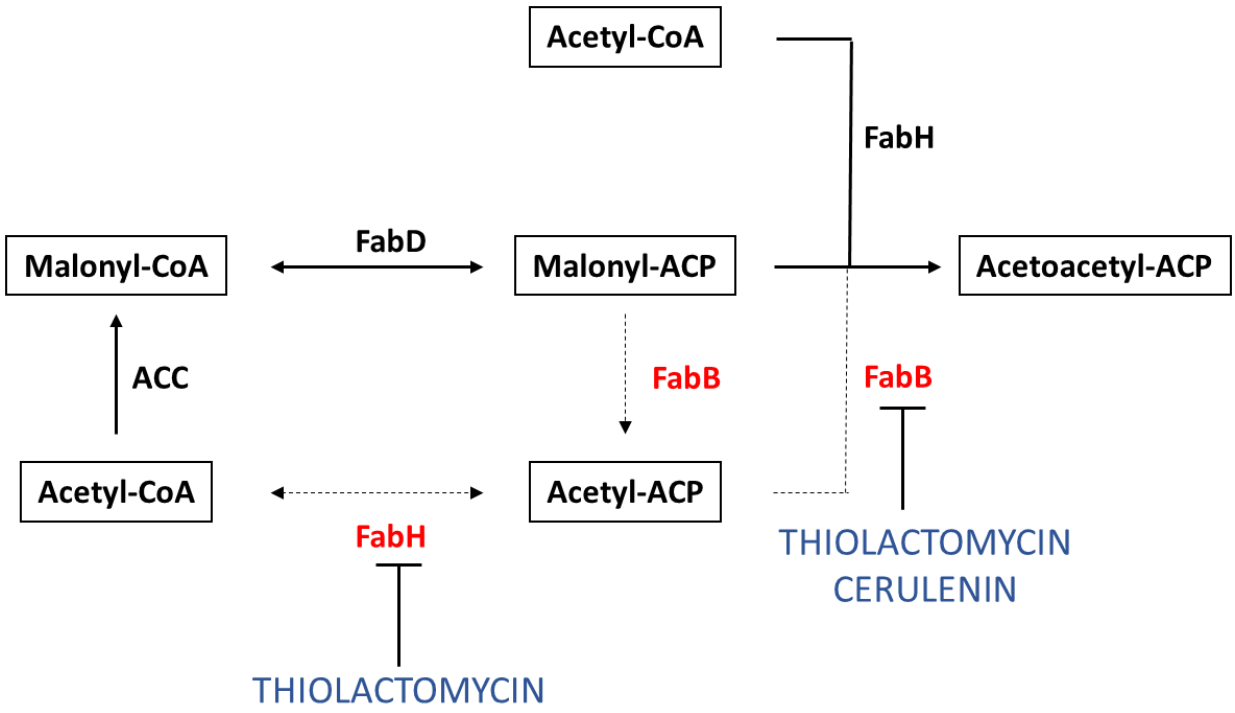

Figure 1.3 Model of the initiation step of fatty acid synthesis. More than 90% of flux to acetoacetyl-ACP is through FabH, which condenses acetyl-CoA with malonyl-ACP. Malonyl-ACP is synthesized by first carboxylating acetyl-CoA using the enzyme acetyl-CoA carboxylase (ACC). This malonyl group is transferred to ACP using the malonyl-CoA-ACP acyltransferase FabD. This is indicated by solid lines. However, the remaining 10% is thought to involve acetyl-ACP which might be produced by decarboxylating malonyl-ACP or by transacylating an acetyl group from acetyl-CoA to holo-ACP. Popular inhibitors and their targets are indicated.

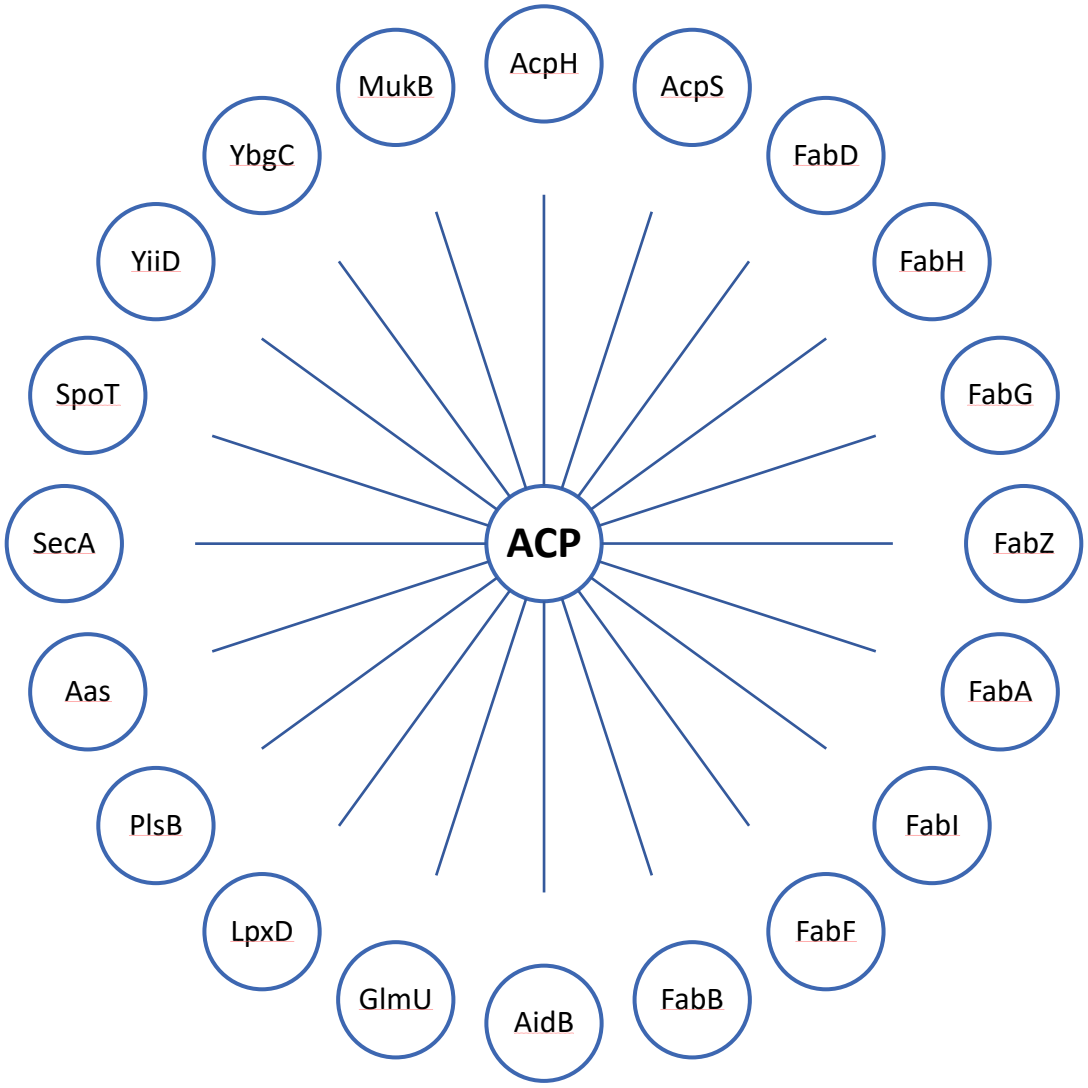

Figure 1.4 Interactome of *E. coli* ACP. A spoke diagram depicting the *E. coli* proteins that are known to interact with ACP as identified by tandem affinity purification. Of these FabD, FabH, Fab Z, FabG, FabA, FabI, FabF and FabB are fatty acid biosynthetic genes and AcpS and AcpH are involved in 4'-phosphopantetheine turn over. ACP seems to interact with proteins in other pathways with no direct relationship to fatty acid synthesis like SpoT (stress response) and SecA (pre-protein translocase).

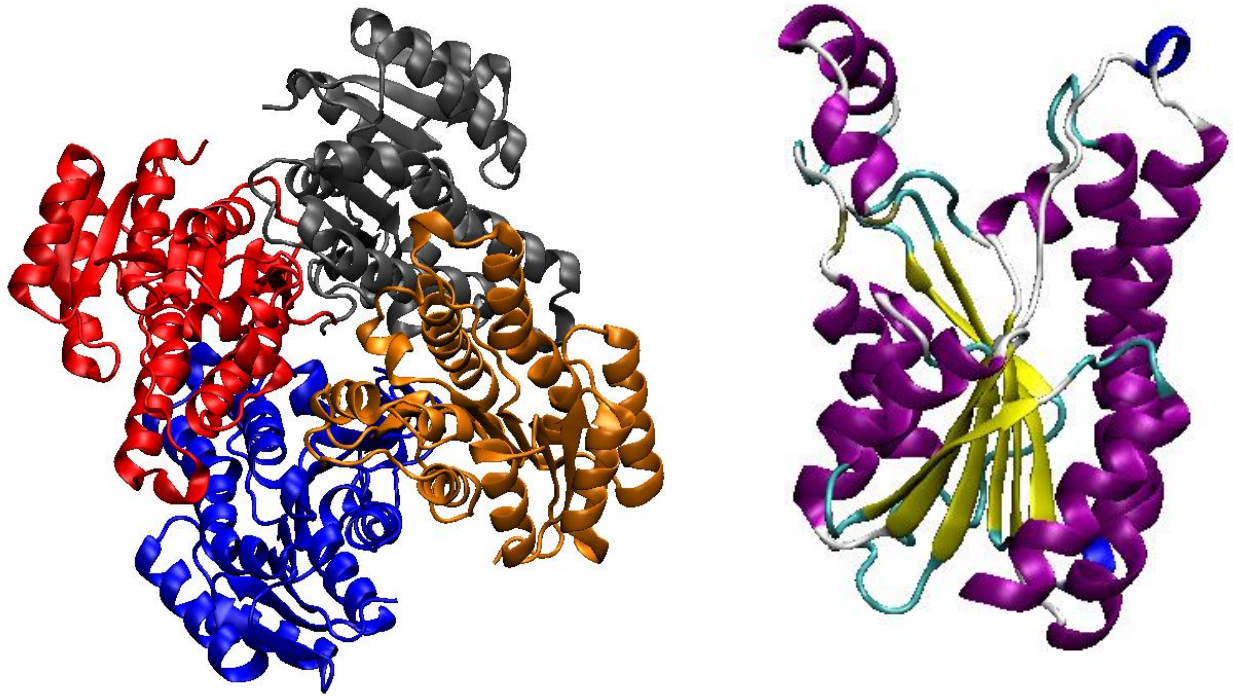

Figure 1.5 Ribbon representation of *E. coli* FabG in its tetrameric and monomeric forms.

The FabG monomer contains a typical Rossmann-fold structure with a twisted parallel beta sheet composed of seven beta strands flanked on both sides by eight alpha helices. Though no data exists about the native oligomerization state of *E. coli* FabG, it is known to function as a tetramer (dimer of dimers). Structures obtained from PDB (1I01) and visualized using VMD software.

Chapter Two

IDENTIFICATION AND CHARACTERIZATION OF FATTY ACID SYNTHESIS INTERMEDIATES IN *ESCHERICHIA COLI* AND *BACILLUS SUBTLIS*

2.1 INTRODUCTION

As described in Chapter I, ACP is the central player in bacterial fatty acid synthesis. Each fatty acid biosynthetic intermediate is covalently attached to an ACP molecule via a thioester bond to the thiol in its prosthetic group, 4'-phosphopantetheine. The prosthetic group comes from the non-nucleotide portion of CoA, and its attachment to ACP is catalyzed by a holo-ACP synthase, AcpS [15]. The prosthetic group may also be cleaved by the action of a phosphodiesterase, AcpH [20]. However, this turnover is independent of fatty acid synthesis where holo-ACP is released only when its acyl group gets transferred to complex lipids. This happens once the acyl group reaches a chain length desired by its target enzymes.

The major destination for long chain fatty acids in the cell is in the structural phospholipids of the cellular membranes. Thus, a bulk of the studies of bacterial fatty acid synthesis have proceeded by analyzing the fatty acid composition of the phospholipids. This was further facilitated by simple and convenient methods of organic extraction like that by Bligh and Dyer [21] and by fast separation techniques like thin-layer, liquid and gas chromatography. However, it is now clear that fatty acid synthesis acts as a feeder pathway for several other essential pathways in bacteria. In *E. coli*, where

this is best studied, biotin and lipoic acid synthesis, lipid A synthesis and quorum sensing amongst others directly depend on shorter acyl-ACP intermediates of fatty acid biosynthesis. In fact, several of the enzymes [13,22,23] in these pathways require the acyl groups to be specifically bound to ACP and not CoA or be in their free-acid form. Further, from several protein interactome studies [19,24,25] undertaken in *E. coli*, it is clear that ACP interacts with enzymes in a variety of processes such as stringent response and protein translocation amongst others that have no obvious co-relation with fatty acid synthesis. A comprehensive understanding of the metabolic flow of acyl species in the cell is unavailable.

In *Bacillus subtilis*, a model organism for Gram-positive bacteria, even less information is available about fatty acid synthesis and its interplay with other dependent pathways. Though the individual enzymes in fatty acid synthesis have been characterized and are akin to those in *E. coli*, *B. subtilis* synthesizes very little saturated and unsaturated fatty acid and instead uses branched-chain fatty acids in its phospholipids. The *B. subtilis* enzymes, thus, have differing configurations to accommodate the branched methyl and methylene groups. Further, while Gram-positive organisms lack an outer membrane and hence lipid A, their other fatty acid synthesis-dependent pathways like biotin and lipoic acid synthesis utilize straight-chain intermediates [26]. Interestingly, certain Gram-positive organisms like *Streptococcus pneumoniae* and *Enterococcus faecalis* can survive at least partially on exogenous fatty acids, and exogenous fatty acid suppress *de novo* fatty acid synthesis. No conclusive evidence is available regarding whether these exogenous fatty acids are incorporated into their phospholipids as ACP thioesters or otherwise. A snapshot of the acyl species under various physiological conditions is necessary to get a holistic view of these bacteria.

This chapter focuses upon the development of methods that can reliably and quickly isolate acyl-ACP species in a highly purified manner that is amenable to an unbiased mass spectrometric detection analysis. Determination of acyl-ACP species has been the subject of study for several decades. Most methods exploit the fact that all synthesized acyl intermediates are covalently linked to ACP by a thioester bond and can be detected as a modification on the ACP. Conveniently, *E. coli* ACP is one of the most abundant monomeric proteins, comprising more than 0.25% of total proteins and accounting to roughly 60,000 molecules per cell. This is unsurprising since, functionally, ACP is a cofactor and not an enzyme. Further, certain unique properties of ACP like its small size (8.8 kDa), unusually low isoelectric point (4.1) and flexible conformation have been repeatedly exploited.

Earlier studies used radio-labeled or antibody-based detection methods followed by resolution polyacrylamide gel electrophoresis (PAGE). Conformation-sensitive PAGE, first developed by Rock and Cronan[27] and subsequently markedly improved by Post-Beittenmiller et al [28] to study spinach acyl-ACP species, is still popular, and separation relies on the varied conformations of different acyl-ACPs. Although, still extremely effective for defined *in vitro* separations, this method requires authentic standards for each species being identified. Moreover, it suffers from poor resolution and limited predictability of the migration pattern of any given acyl-ACP.

With the advent of more sophisticated mass spectrometric methods [29-31], efforts have been made to analyze the acylation pattern via liquid chromatography-mass chromatography under specific scenarios. These methods are extremely effective at identifying expected species but are severely

limited in their capacity to identify novel species. Unbiased approaches to differentiate all fatty acid intermediates suffer from poor reproducibility, require elaborate instrumentation and cannot be used in an ideal setting for determining intracellular fluxes.

2.2 MATERIALS AND METHODS

2.2.1 Bacterial Strains, Plasmids and Materials

The bacterial strains used were derivatives of *E. coli* K-12, *Bacillus subtilis* 168, *Pseudomonas aeruginosa* PAO1, *Enterococcus faecalis* FA2-2, *Clostridium acetobutylicum* ATCC 824 and *Agrobacterium fabrum* C58 (Table 2.1). All strains were grown in LB media. *C. acetobutylicum* was grown anaerobically. The minimal medium for *E. coli* contained M9 salts, 0.4% glucose and 0.1% Casamino acids. General defined media for *B. subtilis* was Spizizen salts[32] consisting of 0.2 % $(NH4)_2SO_4$, 1.4% K_2HPO_4, 0.6% KH_2PO_4, 0.1% sodium citrate · $2H_2O$, 0.02% $MgSO_4 \cdot 7H_2O$] plus trace elements [$MgCl_2 \cdot 6H_2O$, $CaCl_2$, $FeCl_2 \cdot 6H_2O$, $MnCl_2 \cdot 4H_2O$, $ZnCl_2$, $CuCl_2 \cdot 2H_2O$, $CoCl_2.6H_2O$ $NaMoO_4.2H_2O$], 0.4% glucose, and 1 mM $MgSO_4 \cdot 7H_2O$ and 0.01% tryptophan. Antibiotics were used in the following concentrations (in μg/ml): Zeocin (25), streptomycin (100), sodium ampicillin (100) and kanamycin sulfate (50).

Plasmids pKD46, pKD13 and pCP20 were obtained from the *E. coli* Genetic Stock Center (CGSC) at Yale University while plasmid p7Z6 (BGSC ID ECE202) was obtained from the Bacillus Genetics Stock Center. Oligonucleotides were purchased from Integrated DNA Technologies. PCR amplification was performed using Q5 polymerase (New England BioLabs) according to the manufacturer's specifications. DNA constructs were sequenced by ACGT, Inc. Reagents and

chemicals were obtained from Sigma-Aldrich and Fisher Scientific, unless otherwise noted. New England BioLabs supplied restriction enzymes and T4 DNA ligase. StrepTrap HP (5ml) and HiTrap DEAE Sepharose FF were purchased from GE Healthcare. Anti-Strep antibody was purchased from IBA Lifesciences and all western blots were performed as per their protocols.

The *acpP* and *acpA* genes were amplified from the *E. coli* MC1061 and *B. subtilis* 168 genomes by PCR using primer sets *acpP* BamHI F/*acpP* XhoI R and *acpA* BamHI F/*acpA* XhoI R, respectively, with the reverse primers adding a carboxy-terminal Strep-tag II. These amplicons were digested with BamHI and XhoI and ligated with pET28a digested with the same enzymes to give pSW12 and pSW13, respectively.

2.2.2 Construction of Strep-tagged ACP

The *acpP* gene of *E. coli* MC1061 was tagged with a 24-nt strep-tag II sequence (encoding WSHPQFEK) with a 6 nt linker (encoding SA) placed before the stop codon of *acpP*. This was performed using the method of Datsenko and Wanner[33] with PCR primers acpP P1 and acpP P2 used to generate the recombinogenic PCR product with a 24-nt strep-tag II. Recombinants were selected on kanamycin containing solid media and verified by sequencing to contain the Strep-tag at the correct genomic location. This strain was called SW12. and was further transformed by pCP20 to eliminate the antibiotic resistance marker resulting in strain SW13.

The *acpA* gene of *B. subtilis* BSPC10 was tagged in an equivalent manner but using the method of Yan et. al [34] resulting in strain SW14. A fragment containing a Zeocin resistance gene flanked by lox sites was amplified from plasmid p7Z6 using primers acpA F2 and acpA R2. Five hundred

bp upstream and downstream fragments were amplified from the *B. subtilis* 168 genome using primer sets acpA F1/acpA R1 and acpA F3/acpA R3, respectively with the upstream fragment containing a 24-nt strep-tag II sequence (encoding WSHPQFEK) with a 6 nt linker (encoding SA) placed before the stop codon of *acpA.* Strain SW15 was generated by expressing the IPTG-inducible Cre recombinase encoded on the genome, leaving behind only a *lox72* scar on the genome downstream of the strep-tagged *acpA* gene.

2.2.3 Growth Measurements

Cultures were started in LB medium with the necessary supplements and antibiotics. Overnight cultures were diluted to an OD of 0.05 and growth was recorded by making optical density measurements every ten minutes using a Bioscreen C (Growth Curves USA) with rapid shaking for 7 h at 37˚C. All growth curves are plotted as an average of 5 identical samples.

2.2.4 *In vitro* Fatty Acid Synthesis Assay

The integrity of the Strep-tagged ACP was characterized using purified *E. coli* enzymes to carry the enzymes in fatty acid biosynthesis. To produce malonyl-ACP-S Tag, the fatty acid synthesis assay mixtures, which contained 0.1 M sodium phosphate (pH 7.0), 5 mM dithiothreitol, 2 μg holo-ACP-S Tag, and 400 μM malonyl-CoA and *E. coli* FabD (3 μg) in a final volume of 50 μL were incubated at 37°C for 1 h. Then, the following purified His-tagged *E. coli* proteins at 1 μg/assay were added: FabG, FabH, FabA and FabI plus acetyl-CoA (400 μM). NADH (200 μM) and NADPH (200 μM) were added to 30 μl of the previous reaction mixture and the volume was adjusted to 40 μl with H_2O. The reaction mixtures were incubated at 37°C for 1 h. The reactions were stopped by placing on ice and mixed with gel loading buffer and analyzed by conformation-

sensitive gel electrophoresis (CS-PAGE) on 20% polyacrylamide gels containing a 0.5 M urea [28,35]. The gels were stained and visualized using Coomassie Brilliant Blue (CBB).

2.2.5 Radioactive Labeling of Phospholipids and ACP

For labeling phospholipids, *E. coli* cultures were grown at 37°C in LB medium (1% tryptone, 0.5% yeast extract, 1% NaCl) with 5 µCi of sodium [1-^{14}C] acetate until an optical density of 1.0 was attained. The phospholipids were extracted by the method of Bligh and Dyer [21], separated by thin-layer chromatography and detected by phosphorscreen autoradiography using a GE Typhoon FLA7000 phosphorimager. The intensity of the spots was quantified using the Image Quant TL software (GE Healthcare) relative to the wildtype. Radiolabeled fatty acids were purchased from Moravek Biochemicals. All thin-layer plates were scored into lanes to prevent cross-contamination.

Methionine auxotrophs of *E. coli* and *B. subtilis* were grown in 1 ml of LB media with 10 µCi of L-[35S]-methionine added. Growth was allowed to proceed until OD 1.0, and the cells were harvested by centrifugation. Cells were resuspended in 100 µl of 50 mM Tris-HCl buffer pH 8.0. Ten µl was used for measuring protein concentration using the Biuret method[36]. Ten µl of a fresh lysozyme stock (10 mg/ml) was added to the remaining cells, and the mixture was incubated on ice for 15 min. Eight µl of a 500 mM EDTA solution was added, gently mixed and further incubated on ice for 15 min. Bugbuster reagent (Novagen, 12µl of 10X) was added, and lysis was allowed to proceed at room temperature for 30 min. Lysed cultures were centrifuged at 4˚C for 30 mins and their components were separated by CS-PAGE as described before. Gels were stained with CBB for direct visualization, while radioactive incorporation was visualized by

phosphoscreen autoradiography. EasyTag™ L-[35S]-methionine, 500 µCi (18.5MBq) was purchased from PerkinElmer.

ACPs of several bacteria were specifically labelled with ^{14}C-octanoic acid as follows. Bacterial cultures were grown to an OD of 1.0 in LB, and their cells collected by centrifugation. Cells were resuspended in 10 mM Tris-HCl buffer pH 10.0 containing 5 mM dithiothreitol, and they were lysed using multiple passes through a French pressure cell. The lysate was clarified by centrifugation and the cells were stored on ice for 8 h. The lysate was reclarified by centrifugation and was dialyzed overnight against 50 mM Tris-HCl buffer at pH 7.0 using a 1000 MWCO membrane. The resulting lysate was concentrated to roughly 40 mg/ml (measured by the micro-biuret method) and stored at -20˚C for further use. A 50 µl reaction was setup to convert the apo-ACP in the lysate to holo-ACP. The reaction mixture contained 40 µl of the lysate, 1 µg of purified *B. subtilis* Sfp, 10 mM $MgCl_2$, 5 mM lithium CoA, 50 mM Tris pH 8.0 and 1 mM DTT, and the reaction proceeded at 37˚C for 2 h. Purified *Vibrio harveyii* Aas (5µg) was added to this together with 5 µCi of ^{14}C-octanoic acid (Moravek Biochemicals) and further incubated at 37˚C for 4 h. The entirety of this reaction was separated by SDS-PAGE and visualized/quantified by phosphoscreen autoradiography as described earlier.

2.2.6 Purification of Strep-tagged ACP

All procedures were performed at 4°C unless stated otherwise. The starting material for this purification was from cells obtained from 500 ml of *E. coli* or *B. subtilis* cultures at OD 1.0. Cells were resuspended in Buffer A (50 mM ammonium acetate pH 5.5) and lysed using a French pressure cell. For viscous preparations, the lysate was further subjected to sonication at the highest

permissible setting at a 100% duty cycle for 5-10 seconds. The lysate was then clarified by centrifugation at 48,400 x *g* for 30 min. The supernatant was recovered and subjected to a second round of centrifugation.

The supernatant was applied to a 5 ml GE HiTrap DEAE FF column pre-equilibrated with buffer A. Elution fractions between 477 mM and 667 mM (IEX Fraction) were collected from a 50 mM-1000 mM ammonium acetate salt gradient. The IEX fraction was concentrated ten-fold in a 3 MWCO ultrafiltration device and diluted to the original volume in Buffer C (150 mM ammonium acetate, 25 mM ammonium bicarbonate, pH ~7.4). This was then applied to a pre-equilibrated 5 ml GE StrepTrap HP column. The column was further washed with buffer C until no material appeared in the effluent. ACP was selectively eluted using 2 column volumes of Buffer D (buffer C with 2.5 mM desthiobiotin). This was concentrated 20X using a 3 MWCO ultracentrifugation device and lyophilized for storage.

2.2.7 Mass Spectral Analysis

The lyophilized samples were resuspended in 10 mM ammonium bicarbonate and analyzed by direct infusion in to an Orbitrap Fusion Mass spectrometer. The mass spectrum was deconvoluted and analyzed using the Xcalibur 2.0 software. Acyl-modifications were assigned based on mass. The samples were additionally digested with the enzyme Glu-C, and the peptide fragments were subject to LC-MS and peptide mass fingerprinting using the MASCOT server (Matrix Science). This returned a set of post-translational modifications which was used to validate those observed by direct infusion.

2.3 EXPERIMENTAL RESULTS

2.3.1 Carboxy-terminal Strep-tagged ACP is Functional *in vivo*

The genomes of *E. coli* MC1061 and *B. subtilis* 168 were modified such that a Strep-tag II (WSHPQFEK) was introduced at the carboxy-terminus of their respective ACP encoding genes, *acpP* or *acpA* (Fig. 2.1A). The Strep-tag had no observable effect on growth in *E. coli* or *B. subtilis* (Fig. 2.1B). Since *E. coli acpP* is within the fatty acid operon, the effect of the tagging procedure was evaluated by use of the downstream gene *fabF*. A *fabF* mutant is deficient in producing C18:1 fatty acids. While the presence of a kanamycin resistance gene, used as a selection for homologous recombination, significantly reduced cis-vaccenate levels (Fig. 2.2) in phospholipids, elimination of the kanamycin cassette restored wild type levels, indicating that the Strep-tag and downstream FRT scar did not affect *fabF* expression.

2.3.2 Strep-tagged ACP is Easier to Measure and Visualize than its Native Form

The native ACP protein in *E. coli* and *B. subtilis* lacks aromatic residues that are responsible for the typical protein absorbance measured at 280 nm and for binding with dyes like Coomassie brilliant blue. The Strep-tag II introduces a tryptophan residue that can overcome this 280 nm absorbance deficiency. Strep-tagged ACP also stained better in polyacrylamide gels than native ACP, and absorbance measurements led to a more accurate estimation of protein concentration. Further, the availability of a commercial antibody against the Strep-tag allows for easy visualization of Strep-tagged ACP from complex cellular lysates through western blotting (Fig. 2.3A).

2.3.3 Strep-tagged ACPs are Functional *in vitro*

Strep-tagged ACP from E. coli (EcACP-STag) and *B. subtilis* (BsACP-STag) were obtained as a mixture of their apo and holo forms when overexpressed in and purified from *E. coli* BL21(Star). This indicates that the native *E. coli* AcpS can function on EcACP-STag and BsACP-STag *in vivo*. Both proteins could be completely converted to their holo form by the action of Sfp which could be separated on a CS-PAGE from the apo form. The purified ACP migrated at a slightly different distance on a urea containing polyacrylamide gel than their native forms. Holo-EcACP-STag was also used in an *in vitro* reconstituted fatty acid system wherein it was first converted to malonyl-ACP using FabD and malonyl-CoA and finally extended to butyryl-ACP by the action of purified *E. coli* FabH, FabG, FabA and FabI (Fig. 2.3B). A similar result was observed with the BsACP-STag protein (not shown). This demonstrates that the tag does not block the activity of typical fatty acid enzymes.

2.3.4 *E. coli* has Five to Ten-fold more ACP than *B. subtilis*

Western blots of cellular lysates of SW13 and SW15, using antibodies targeting the Strep-tag, show much lower levels of ACP in *B. subtilis* in comparison to *E. coli*. The anti-Strep antibody however, can recognize both Strep-tagged ACPs equally (Fig 2.4). Densitometric calculations place ACP levels in *B. subtilis* to be around 8% of that in *E. coli*, when normalized against total cellular protein content. This was further confirmed by ^{35}S-methionine labeling of total proteins in a *B. subtilis* 168 Δ*metE* strain and an *E. coli* BW25113 Δ*metE* strain and separating ACP from other cellular proteins by SDS-PAGE. Abundance counts of *E. coli* and *B. subtilis* ACP obtained by phosphorscreen autoradiography show that *B. subtilis* has roughly 18% of the ACP levels as those in *E. coli*. ACP from cellular lysates of *E. coli* MG1655, *B. subtilis* 168, *Pseudomonas*

aeruginosa PAO1, *Clostridium acetobutylicum* ATCC 824, *Enterococcus faecalis* FA2-2 and *Agrobacterium fabrum* C58 were also specifically labeled with ^{14}C-octanoic acid. This places *B. subtilis* ACP concentrations at 12% of that of *E. coli* (Fig. 2.4B). Surprisingly, similar ACP levels were also observed in *P. aeruginosa* (6%) and *A. fabrum (18%).* Interestingly, almost no signal was observed in *E. faecalis* and *C. acetobutylicum* lysates, indicating a lack of recognition by Sfp of these ACPs or a lack of sensitivity for very low ACP levels by this method (Fig. 2.4C). To see if this difference in protein levels is due to transcriptional regulation, a quantitative RT-PCR was performed with protein normalized RNA isolated from *E. coli* MC1061 and *B. subtilis* 168. The *acpP* mRNA levels in *E. coli* were roughly twice that of *acpA* from *B. subtilis*

2.3.5 The ACP Interactome of *B. subtilis*

Studies in *E. coli* have shown that ACP is extremely interactive with other proteins. In *B. subtilis* we observe that ACP interacts with a large number of proteins (Table 2.4). Most of these proteins play a direct role in fatty acid biosynthesis and include FabF, FabH1, FabH2, FabG and FabI (Fig. 2.5B). FabZ was identified after the affinity purification but not after size exclusion chromatography. Other enzymes such as the different subunits of the pyruvate dehydrogenase complex, ribosomal proteins, sugar transport related proteins and certain unknown proteins are also associated with ACP. The veracity of these interactions has not been verified.

2.3.6 FabF Tightly Associates with *B. subtilis* ACP in a Stoichiometric Fashion

A contaminating band was consistently found in purified *B. subtilis* ACP preparations. These were often identified to be cold shock proteins that are similar in size. However, one ‘contaminant’ persisted through multiple purification columns and was identified to be *B. subtilis* FabF by peptide

mass fingerprinting. In fact, when ACP was purified directly from the streptactin column, FabF was co-purified in a 1:2 ratio. This is expected because FabF is a dimer. In fact, when incubated *in vitro* purified FabF and ACP were observed to migrate as a single peak during gel filtration (Fig. 2.5A). Attempts to crystallize this complex have thus far failed. A similar interaction between *E. coli* FabF and ACP was not observed.

2.3.7 Purification of Acyl-ACP species from *E. coli* and *B. subtilis*

An ammonium acetate buffer at pH 5.5 was chosen to lyse the cells grown in the condition to be tested. Malonyl-ACP, prepared *in vitro,* was found to be stable in this buffer when tested after a 24 h incubation at 4˚C, corresponding to the average length of purification. The lysate was clarified by ion-exchange chromatography using a DEAE Sepharose column. The chromatography conditions were systematically tested in *E. coli* to ensure complete recovery of acyl-ACPs while eliminating as many contaminating proteins as possible. For *E. coli,* the optimal window for elution was between 450 mM and 650 mM ammonium acetate. *B. subtilis* required a more relaxed window of elution between 350 mM and 650 mM ammonium acetate (Fig. 2.7A). Complete recovery was confirmed by monitoring all fractions for ACP by CS-PAGE (Fig. 2.7B). However, a significant level of contaminating protein was still seen after this step.

Binding to the streptactin column requires a pH above 7. To ensure optimum binding without thioester hydrolysis, the stability of pure malonyl-ACP was tested in buffers containing various proportions of ammonium acetate and ammonium bicarbonate. Malonyl-ACP was found to be stable for at least 12 h (maximum time tested) in a 150 mM ammonium acetate and 25 mM ammonium bicarbonate buffer that has a pH of ~ 7.4. To minimize handling time at the higher pH,

the eluate from the ion-exchange column was concentrated ten-fold in a 3 kDa spin column, diluted in the above-mentioned buffer to the original volume and directly injected into an equilibrated streptactin column. Acyl-ACPs were eluted with 2.5 mM desthiobiotin after washing the column until a flat baseline was achieved (Fig. 2.7A). The eluate was free of contaminants and gave a single band when visualized by CS-PAGE (Fig. 2.7B). This was also confirmed by peptide mass fingerprinting. The purified acyl-ACPs were immediately lyophilized for further analysis.

2.3.8 Mass Spectrometric Analysis and Validation

The lyophilized samples were dissolved in 10 mM ammonium bicarbonate buffer with formic acid and directly injected into an Orbitrap-based mass spectrometer. Acyl-ACP identifications were made based on mass. In *E. coli*, very little or no apo-ACP was observed. The major species observed at every stage of growth were, however, acetyl ACP. Some malonyl-ACP was detected as well. Acetyl-ACP remains the major species in *E. coli* from early exponential growth until stationary growth. Longer chain saturated fatty acids (14-18 carbon length) were mainly observed earlier during growth with shorter chain accumulating as growth progressed. The levels of malonyl-ACP remained stable regardless of growth phase except during the stationary phase wherein it diminishes by 60-70% of earlier levels. Interestingly, during early growth, *E. coli* accumulated some decanoyl-ACP (16%) which fell to 5% or less as exponential growth continued into stationary phase.

In *B. subtilis*, acyl-ACP yields are much lower, but the signal/noise ratio observed during mass spectrometry remains the same. In fact, the major species of *B. subtilis* is also acetyl-ACP. Other acyl-ACP species were also reliably observed.

2.3.9 The Metabolic Role of Acetyl-ACP

Acetyl-ACP remains the predominant species in both bacteria. In *E. coli*, acetyl-ACP is observed across its various stages of growth in rich media. Purified malonyl-ACP has been shown to be stable under purification conditions. Acetyl-ACP is observed in *E. coli* even if the culture is instantly lysed using 10% TCA followed by an acetone wash before extraction. This suggests that acetyl-ACP is an authentic product of metabolism present in the cell. The source and role of acetyl-ACP is still being investigated.

2.4 DISCUSSION

Fatty acid synthesis is a vital pathway in bacteria which primarily feeds phospholipid synthesis along with other dependent pathways like vitamin synthesis, lipid A synthesis and quorum sensing amongst others. The intermediates of this pathway are covalently liked to ACP via a thioester bond. Even though ACP is easily purified due to its small size and acidic nature, preservation of the thioester bond can be difficult under *in vitro* conditions.

In this Chapter, I describe a method for isolating acyl-ACPs in an extremely pure manner while maintaining the thioester bond and the ensuing acyl modifications of ACP. This required the introduction of an inert Strep-tag onto the ACP protein, which could be then be isolated using a Streptactin column that can specifically recognize the Strep-tag with high affinity but be eluted with a small non-reactive molecule, desthiobiotin (while desthiobiotin can be removed from the Streptactin column under alkaline conditions, biotin binds it irreversibly). Hence, to clarify the lysate while removing biotin present in it, the samples were first subjected to anion-exchange

chromatography. Previous attempts to use a strong ion-exchanger resulted in increased contamination with small RNA fragments (Table 2.5) that possess a similar acidic charge. Hence, a weak ion-exchanger like DEAE was preferred. New variants of Streptactin, called Streptactin XT, are now available that can be directly eluted with biotin.

A volatile buffer system (ammonium acetate:acetic acid) was chosen with a pH of 5.5 to preserve the thioester bond. However, the biggest and only limitation of using the Strep-tag for this particular purpose is its requirement for a pH above 6.5 to bind to the streptactin column. Hence, the pH was adjusted to 7.5 using ammonium bicarbonate after ion exchange chromatography. Other affinity tags like the hexahistidine tag also have this requirement. In addition, the imidazole eluant would hydrolyze the thioester and is hence avoided (Table 2.5). The low pH during the initial step served to reduce contaminant binding during anion-exchange and the pH shift after that step additionally destabilized other contaminant proteins that had an isoelectric point within this range. Ammonium acetate was preferred over buffers with sodium and potassium ions to avoid the formation of stable salt bridges with the ACP species. Though, Na^+ and K^+ can improve ionization during mass spectrometry, salt bridges can severely convolute peak assignment during identification of novel species.

E. coli ACP is an extremely abundant protein with 60,000 copies per cell. However, such data are unavailable in other organisms. Our data indicate that *B. subtilis* consistently has five to ten-fold lower total ACP levels than *E. coli*. The true value is probably closer to ten-fold, since cold-shock proteins expressed during handling are a major contaminant of ACP preparations in *B. subtilis* and run with the same mobility as ACP in CS-PAGE. Our preliminary evidence using ^{14}C-octanoic

acid labelling of the ACP thiol shows that this might be the case in several other Gram-positive and Gram-negative organisms. In fact, we observe little to no ACP in *Clostridium acetobutylicum* and *Enterococcus faecalis* when normalized to total protein content. ACP is found to associate strongly with FabF both *in vivo* and *in vitro*. However, the reciprocal experiment wherein FabF is tagged with a Strep-tag at its carboxy terminus on the genome and ACP is expected to co-elute, has proved difficult. The tag at the C-terminus remains inaccessible to the column even with the use of longer linkers, and almost no FabF can be isolated from *B. subtilis.* One speculation is that ACP binding to FabF results in the tag being buried in the protein-protein interface. *B. subtilis* ACP can also be reliably pulled down with the other fatty acid synthesis proteins. In addition, proteins like the biotin carboxylase carrier protein and pyruvate dehydrogenase were also isolated, suggesting that some indirect interactions were also captured. Our method to identify acyl-ACPs was also effective in *B. subtilis*, regardless of its lower ACP levels, suggesting that this methodology can be scaled while retaining its sensitivity.

Previous reports suggested that the majority of ACP in *E. coli* exists in its holo form with only a small population being acylated [37]. These studies used CS-PAGEs to make these estimates, and acetyl-ACP migrates similar to holo-ACP on a CS-PAGE under most conditions. On the contrary, we observe that a majority of the ACP in *E. coli* exists as acetyl-ACP. Acetyl-ACP was suggested to be the original primer for fatty acid initiation. This theory had fallen out of favor when it was demonstrated in in vitro that FabH can only function with acetyl-CoA, with more than 90% of the fatty acid flux flowing through this enzyme. 10% of the total enzymatic activity of FabH is a transacylase activity that can convert acetyl-CoA to acetyl-ACP. Also, malonyl-ACP could spontaneously decarboxylate *in vivo* to give acetyl-ACP. However, *fabH* can be knocked out in *E.*

coli with no significant defect, and malonyl-ACP is an expensive species to accumulate since the generation of each malonyl moiety requires one ATP molecule. The metabolic role and the source of acetyl-ACP and the true initiator of fatty acid synthesis is yet to be characterized.

Similar observations of acetyl-ACP are also made in *B. subtilis*, suggesting a commonality between *E. coli* and *B. subtilis* in this regard. Bacillus species produce very little straight-chain fatty acids. These straight-chain fatty acids might only be required for associated pathways like biotin synthesis and lipoic acid synthesis, but these pathways also maintain other redundant mechanisms[26,38-40]. A study reports the creation of a double knockout of *fabHA* and *fabHB* in *B. subtilis* that can be stably grown in the presence of long chain branched fatty acids[41]. This strain is reported to still incorporate straight chain fatty acids in its membrane, suggesting that a novel initiation mechanism might indeed exist. Our method opens up the possibility of examining bacteria like *E. coli* and *B. subtilis* under new light. With minor genomic modifications, these bacteria can be tested under numerous external conditions including the effect of novel antibiotics, with the method itself being agnostic to the treatment. Additionally, better metabolic models can be built with these data sets, possibly revealing novel modes of regulation and improving their utility to the scientific, clinical and industrial community.

2.5 TABLES AND FIGURES

Strains	Relevant Genotype	Source
MC1061	Wild type strain	Casadaban, et al., 1980
DH5α	*endA1 recA1 0 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17*	Hanahan, et al., 1985
MG1655	*K-12 F– λ– ilvG– rfb-50 rph-1*	CGSC #6300
BL21 (Star)	*ompT hsdSB (rB-, mB-) gal dcm rne131 (DE3)*	Invitrogen
SW12	MC1061 *acpP*-StrepTag::FRT-*KanR*-FRT	This work
SW13	MC1061 *acpP*-StrepTag::FRT	This work
SW14	BS168 *acpA*-StrepTag::*lox69-StrR-lox71*	This work
SW15	BS168 *acpA*-StrepTag::*lox72*	This work
JW3805	BW25113 ΔmetE::Kan	CGSC# 10758
BKK13180	BS168 ΔmetE::Kan	BGSC
BS168	*B. subtilis* wildtype strain; *trpC* auxotroph	BGSC
BSPC10	BS168 ΔamyE::pSpac-Cre $Spec^R$	BGSC
BW25113	*lacI+rrnBT14 ΔlacZWJ16 hsdR514 ΔaraBADAH33 ΔrhaBADLD78 rph-1*	CGSC #7636

Table 2.1 Bacterial Strains used in this study

Plasmids	Relevant Characteristics	Source
pBAD24	Amp^R, Arabinose-inducible expression vector	Guzman, et al., 1995
pET28b	Kan^R, T7 promoter-based, IPTG inducible expression vector	Novagen
pSW04	Kan^R, N-terminal hexahistidine-tagged *E. coli fabG* gene in pET28a	Srinivas et. al, 2017
pSW05	Kan^R, N-terminal hexahistidine-tagged *E. coli fabD* gene in pET28a	Srinivas et. al, 2017
pSW06	Kan^R, N-terminal hexahistidine-tagged *E. coli fabH* gene in pET28a	Srinivas et. al, 2017
pSW07	Kan^R, N-terminal hexahistidine-tagged *E. coli fabA* gene in pET28a	Srinivas et. al, 2017
pSW08	Kan^R, N-terminal hexahistidine-tagged *E. coli fabI* gene in pET28a	Srinivas et. al, 2017
pSW12	Kan^R; PCR amplified Ec *acpP* with C-terminal Strep-tag in BamHI/XhoI sites of pET28a	This work
pSW13	Kan^R; PCR amplified Bs *acpA* with C-terminal Strep-tag in BamHI/XhoI sites of pET28a	This work
pKD46	Amp^R, Ts ori, Arabinose-inducible expression of λ-Red genes	Datsenko and Wanner, 2000
pKD13	Pir dependent plasmid that has FRT-Kan^R-FRT cassette	Datsenko and Wanner, 2000
P7Z6	AmpR; plasmid carrier lox69-Zeo^R-lox71 cassette	Yan et. al, 2008

Table 2.2 Plasmids used in this study

Oligonucleotides	Sequence
acpP P2	AAATCACCACCGTTCAGGCTGCCATTGATTACATCAACGGCCA CCAGGCGTCAGCGTGGAGCCACCCGCAGTTCGAAAAATAAATT CCGGGGATCCGTCGACC
acpP P1	GACAAAAAGATAAAACTCAGGCGGTCGAACGACCGCCTGGAG ATGTTCACTGTAGGCTGGAGCTGCTTCG
acpA F1	TTTATCTCAACTGATATGACAGATAAACTTG
acpA R1	AATCTCTAGAGGATCCCCGGGTACCGAGCTCTTATTTTTCGAAC TGCGGGTGGCTCCACGCTGATTGCTGGTTTTGTATGTAGTTCAC
acpA F2	GAGCTCGGTACCCGGGGATCCT
acpA R2	GCTTGCATGCCTGCAGGTCGAC
acpA F3	GTAGAATCGTCGACCTGCAGGC GCTGATG CTAAAAGTCCCG
acpA R3	CGCCTTTACCCAACAGGA
acpP BamHI F	ACGTCAGGATCCATGAGCACTATCGAAGAACG
acpP XhoI R	GCTACACTCGAGTTATTTTTCGAACTGCGGGTGGCTCCACGCTG ACGCCTGGTGGCCGTTGAT
acpA BamHI F	ACGTCAGGATCCATGGCAGACACATTAGAGC
acpA XhoI R	GCTACACTCGAGTTATTTTTCGAACTGCGGGTGGCTCCACGCTG ATTGCTGGTTTTGTATGTAGTTCAC
Ec AcpP RT F	CTGGTAATGGCTCTGGAAGAAG
Ec AcpP RT R	CTGGTGGCCGTTGATGTAAT
Bs AcpA RT F	GCAGACACATTAGAGCGTGTA
Bs AcpA RT R	CATCTAGGGAATCAGCACCTAAG

Table 2.3 Oligonucleotides used in this study

Protein name	Function	Pathway	Counts
FabF	β-ketoacyl-ACP synthase II	Fatty acid synthesis	16638
ACP	**Acyl carrier protein**	**Fatty acid synthesis**	**8562**
FabH1	β-ketoacyl-ACP synthase III	Fatty acid synthesis	630
FabG	β-ketoacyl-ACP reductase	Fatty acid synthesis	532
PdhC	Pyruvate dehydrogenase E2 subunit	Lipoic acid synthesis	476
PdhB	Pyruvate dehydrogenase E1β subunit	Lipoic acid synthesis	397
HepS	Heptaprenyl diphosphate synthase	Isoprenoid synthesis	397
DegU	Two-component response regulator	Catabolite repression by glucose	324
PdhA	Pyruvate dehydrogenase E1α subunit	Lipoic acid synthesis	321
CarA	Carbamoyl-phosphate synthase	Pyrimidine biosynthesis	292
PtsI	Phosphoenolpyruvate—protein phosphotransferase	Sugar transport	240
RL7	50S Ribosomal subunit L7	Protein synthesis	231
FabH2	β-ketoacyl-ACP synthase III	Fatty acid synthesis	212
GlyA	Serine hydroxymethyltransferase	Purine synthesis	177
PdhD	Dihydrolipoyl dehydrogenase	Lipoic acid synthesis	175
BCCP	Biotin carboxyl carrier protein	Fatty acid synthesis	173
EftU	Elongation factor Tu 1	Protein synthesis	152
YdbB	Putative Mn-binding protein	Unknown function	125
CspC	Cold shock-like protein	Cold adaptation	121
FabI	Enoyl-ACP reductase	Fatty acid synthesis	111
MurB	UDP-N-acetylenol-pyruvoylglucosamine reductase	Cell wall synthesis	109
YhfS	Putative acetyl-CoA acetyltransferase	Fatty acid metabolism	107
RS2	30S Ribosomal protein S2	Protein synthesis	88
Eno	Enolase	Glycolysis	84

Table 2.4 List of proteins interacting with ACP in *B. subtilis.*

Purification Condition	Concerns
Ion-exchange counterions	Contamination by RNA fragments with strong counterions
Buffer pH	pH above 7.5 can cause thioester hydrolysis
Buffer composition	Salts like imidazole can cause thioester hydrolysis
Buffer ions	Salt adduct formation can convolute spectral analysis
Size exclusion	High contaminant levels; Increase in signal/noise ratio during MS

Table 2.5 Considerations for acyl-ACP purification

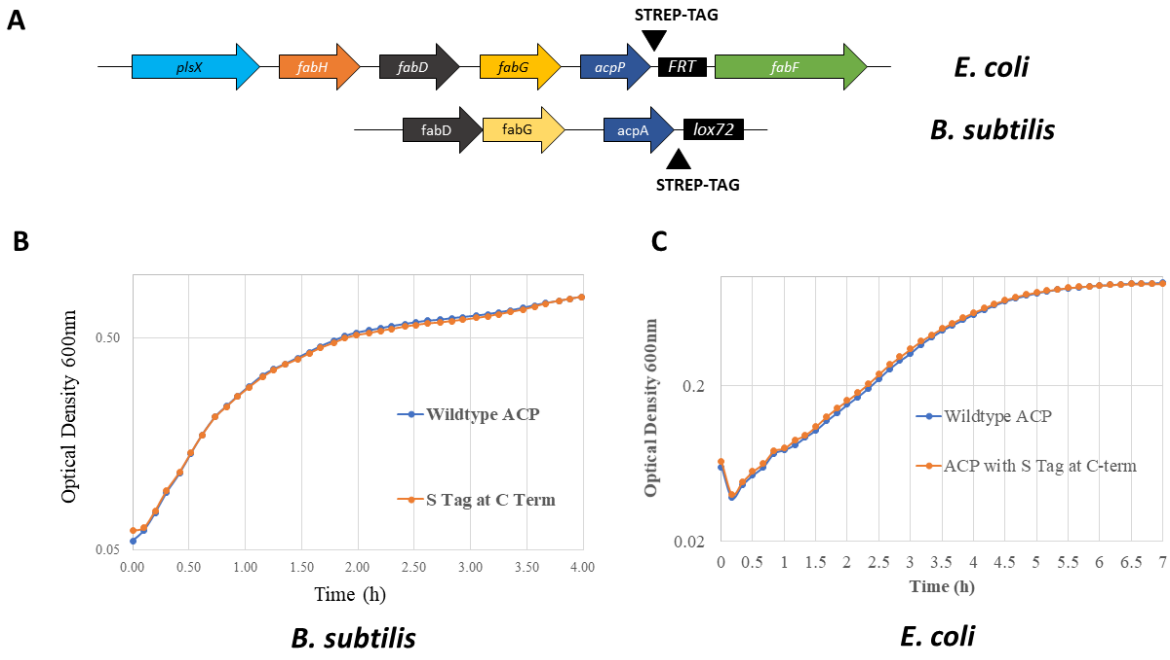

Figure 2.1 Effect of Strep-tag on ACP function in vivo. A) The fatty acid biosynthetic operons of *E. coli* and *B. subtilis* are visualized here with the scar left from the tagging methodology indicated as a black box. The *E. coli* operon has a FRT site between *acpP* and *fabF*. *B. subtilis* has a *lox72* scar downstream of *acpA*. B) Growth curve of wildtype *B. subtilis 168* compared to that with a Strep-tag at the carboxy terminus of ACP C) Growth curve of wildtype *E. coli MC1061* compared to that with a Strep-tag at the carboxy terminus of ACP.

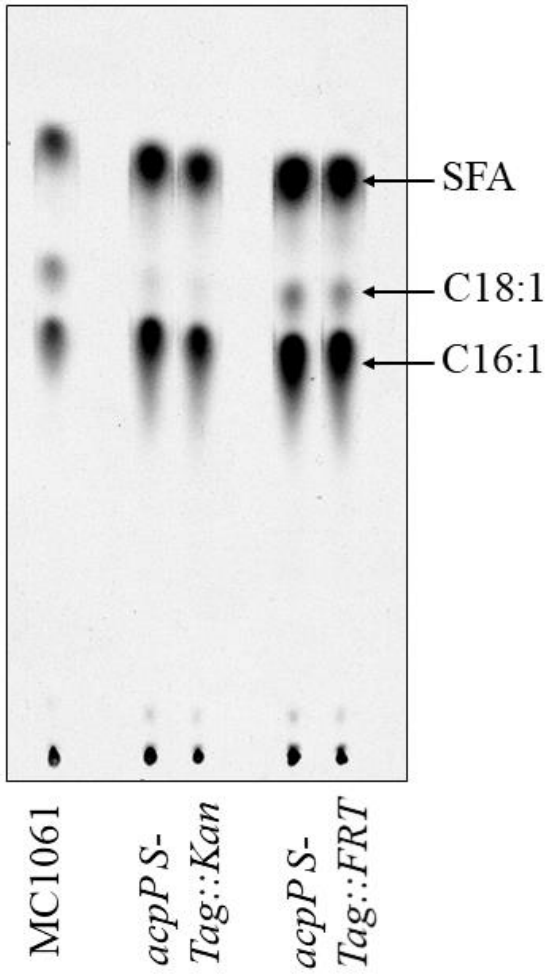

Figure 2.2 *acpP* tagging in *E. coli* has no polarity on downstream genes. An autoradiograph of fatty acid methyl esters from phospholipids isolated from strains grown in ^{14}C-sodium acetate and separated by Ag-thin layer chromatography. Wildtype E. coli MC1061 has saturated fatty acids (SFA), cis-vaccenic acid (C18:1) and palmitoleic acid (C16:1), while MC1061 with a kanamycin resistance cassette between the tagged *acpP* gene and *fabF* is polar (Lanes 2 and 3). However, removal of this cassette results in wildtype fatty acid composition (Lanes 4 and 5).

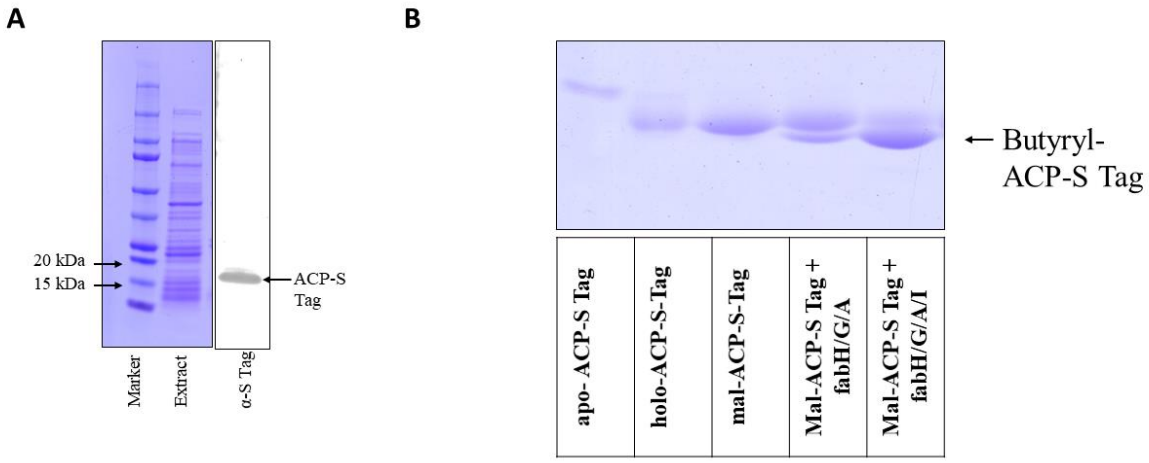

Figure 2.3 Strep-tagged ACP is functional *in vitro.* A) SDS-PAGE and western blot detection of ACP from *E. coli* with Strep-tagged ACP. Strep-tagged ACP runs with an apparent molecular weight around 17.5 kDa due to its acidic nature. B) Apo-, holo- and malonyl-ACP visualized by conformationally-sensitive PAGE (Lanes 1-3). Malonyl-ACP-S Tag along with acetyl-CoA is readily utilized by FabH, FabG, FabA and FabI to give butyryl-ACP-S Tag.

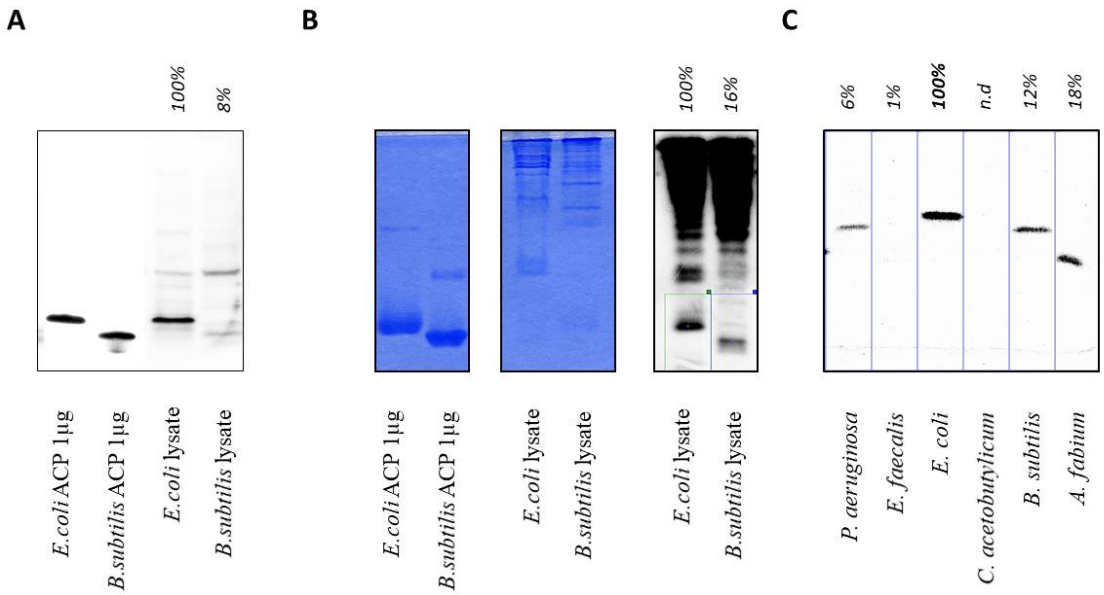

Figure 2.4 Relative amounts of ACP in *E. coli* and *B. subtilis.* A) Western blot using a Strep-tag-specific antibody in cell lysates of *E. coli* and *B. subtilis.* Purified proteins are loaded as controls for antibody specificity. B) ^{35}S-labeled total proteins separated by CS-PAGE. Purified Strep-tagged ACP from *E. coli* and *B. subtilis* is loaded on the same gel to calculate migration distance. Labeled bands corresponding to ACP are quantified using phosphorscreen autoradiography. C) ^{14}C-Octanoate labeled ACP from cellular lysates of multiple bacteria. Quantifications were done using phosphorscreen autoradiography and are indicated above each lane relative to *E. coli* (set to 100%)

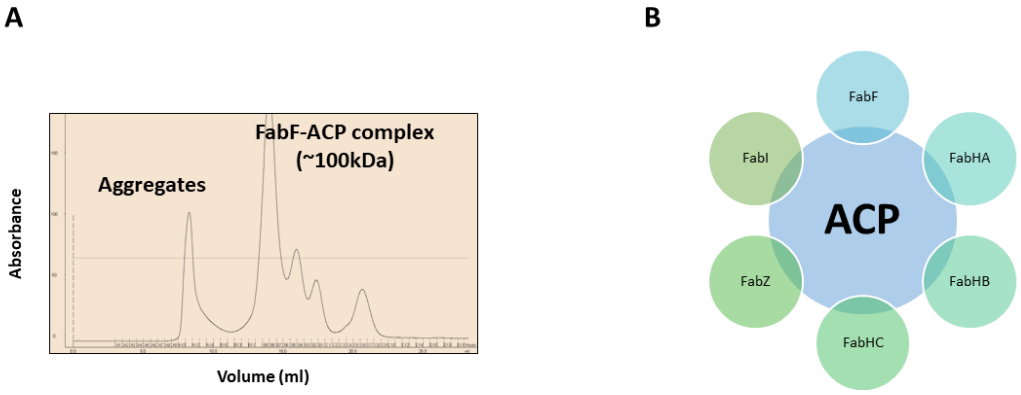

Figure 2.5 *B. subtilis* ACP interacts tightly with some of its partners. A) Size Exclusion Chromatogram of *in vitro* co-incubation of FabF and ACP. FabF and ACP elute as one large peak with a size corresponding to a dimer of FabF and one unit of ACP. B) A schematic of proteins involved in fatty acid biosynthesis that co-elute with *B. subtilis* ACP. A loose complex is proposed, to compensate for lower ACP levels.

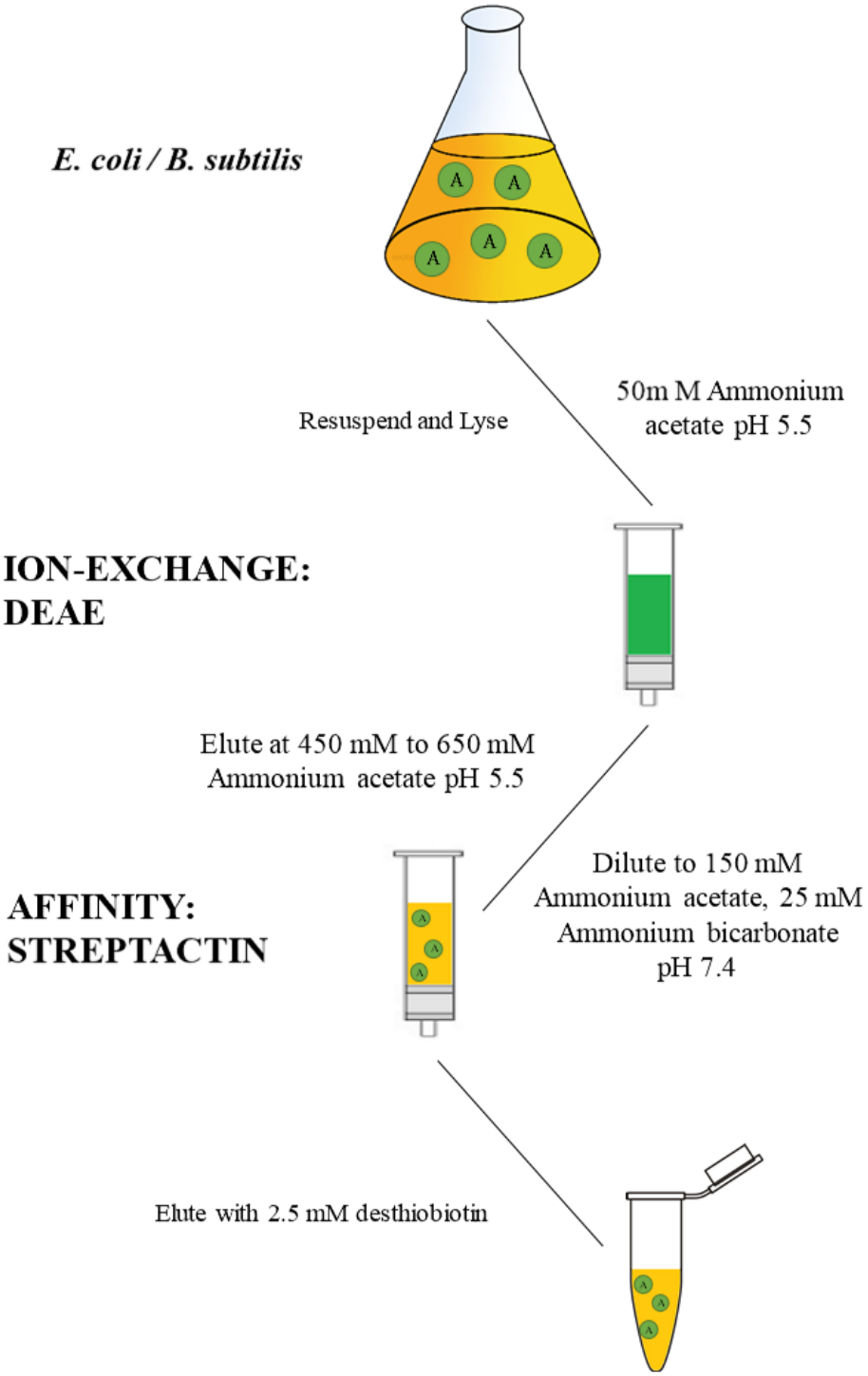

Figure 2.6 Schematic of acyl-ACP purification with final buffer conditions indicated alongside the chromatographic techniques utilized.

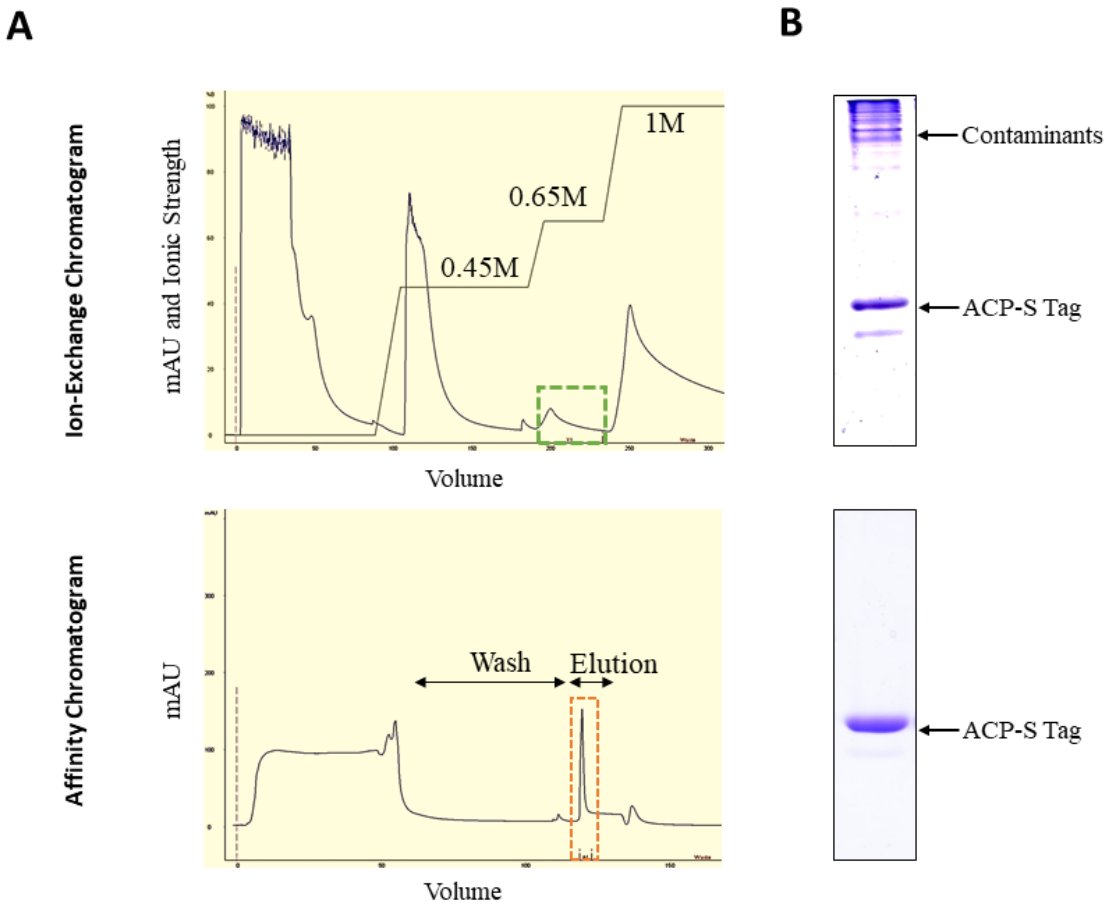

Figure 2.7 Efficiency of the acyl-ACP purification process. A) Chromatographic profiles of ion-exchange chromatography and affinity chromatography with the desired fractions indicated with a dotted box. B) Desired fraction from each step visualized by CS-PAGE. Contaminants are retained near the origin while ACP migrates faster and is indicated.

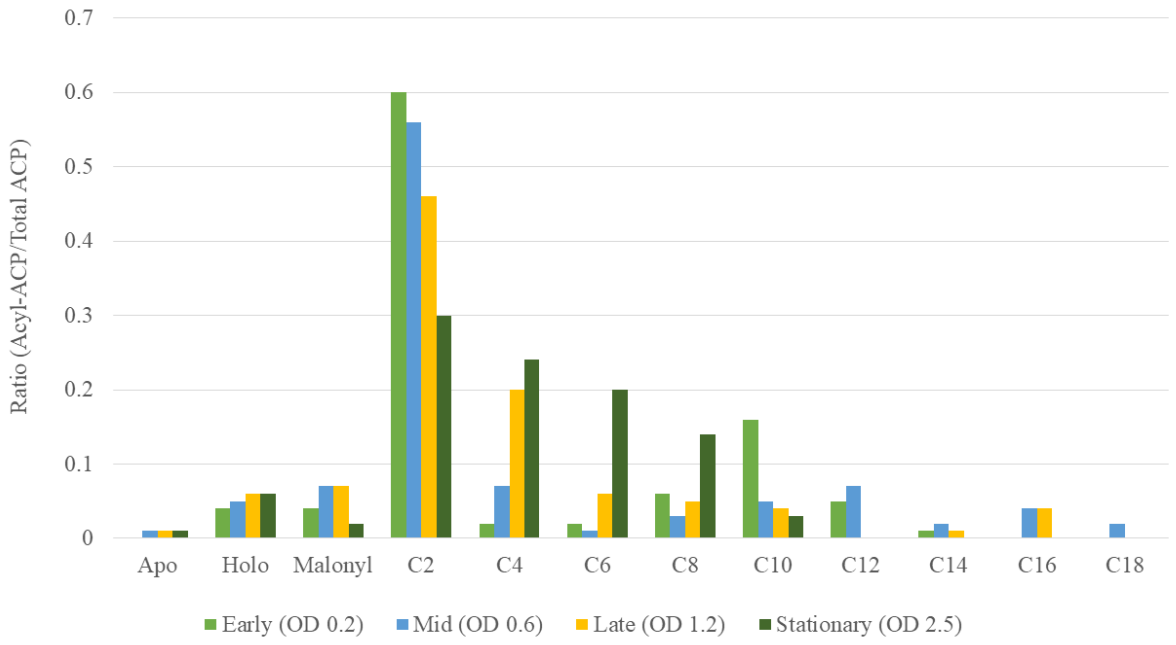

Figure 2.8 A snapshot of acyl-ACPs identified during various stages of growth in *E. coli.* A graph depicting the ratio of apo-, holo- and acyl-ACPs to the total ACP detected during mass spectrometry. The data above only includes saturated acyl-ACPs for simplicity.

Chapter Three

AN EIGHT RESIDUE DELETION IN *ESCHERICHIA COLI* FABG CAUSES TEMPERATURE-SENSITIVE GROWTH AND LIPID SYNTHESIS PLUS RESISTANCE TO THE CALMODULIN INHIBITOR, TRIFLUOPERAZINE

3.1 INTRODUCTION

As discussed in Chapter 1, bacteria and plants utilize type II fatty acid synthesis (FAS), a dissociated system of individual enzymes, to generate the entire fatty acid repertoire from acetyl-CoA. The intermediates are elongated by a series of four enzymes while being covalently bound to the acyl carrier protein (ACP) [8,42] via a thioester linkage. In *Escherichia coli*, a ß-keto acyl-ACP synthase (FabH, FabB or FabF) catalyzes a Claisen elongation reaction, that converts a growing acyl-ACP to its ß-ketoacyl form. The resulting ß-keto group is then reduced by a ß-ketoacyl-ACP reductase (FabG) [43] in a NADPH-dependent manner to a ß-hydroxyl group. The resulting intermediate is dehydrated by a ß-hydroxyacyl-ACP dehydratase (FabA or FabZ) to an enoyl-ACP and finally reduced again by an enoyl-ACP reductase (FabI), in a NADH-dependent reaction, to give an acyl-ACP that is two carbons longer and poised to re-enter the cycle until the appropriate chain length for complex lipid synthesis is achieved. Some of the early intermediates are siphoned off by FAS-dependent synthetic pathways such as those for biotin [44] and lipoic acid [44], whereas longer fatty acids are predominantly utilized for the synthesis of the membrane components, lipid A synthesis [45] and phospholipids [45].

FabG is a member of the short-chain dehydrogenase/reductase enzyme family (SDR), which carry out a variety of dehydrogenation and reduction reactions using a nucleotide cofactor [46]. *E. coli* FabG is functionally active as a dimer of dimers with structural studies suggesting allosteric regulation induced by conformation changes within the tetramer [47]. Though FabG is an essential enzyme, like most fatty acid synthesis enzymes, it is highly conserved across species and is the only isoform required for all ß-keto acyl-ACP reductions in the FAS pathway [48]. An earlier study establishing the role of *fabG* in FAS isolated several temperature-sensitive (Ts) point mutants in *E. coli* (E233K) and *Salmonella enterica* (S224F; M125I and A223T) by a tritium suicide selection [49]. Most reported Ts mutants, produce proteins with missense mutations predicted by computational modeling [50] to cause large changes in denaturation temperatures. In the case of FabG, the Ts mutants produce extremely unstable and thermolabile proteins that very rarely survive purification and enzymatic characterization, even when produced and assayed at permissive temperatures [49].

In this chapter, I report the detection, construction and biochemical characterization of *fabG*Δ*8* which encodes a FabG that lacks eight residues that form part of an important helix-turn-helix motif that participates in dimerization, allostery and formation of the entrance to the active-site tunnel. Only a handful of Ts mutations have been isolated that are caused by partial deletions and the Ts phenotype is usually the result of impaired subunit interactions [51] or decreased protein stability at higher temperatures [52]. Understanding the mechanism of temperature-sensitivity in FabGΔ*8* increases our knowledge of intra-subunit interactions in the FabG tetramer and their contribution to membrane lipid synthesis.

In addition, I also demonstrate the suppressive influence of FabG in a lipid A biosynthetic mutant (*lpxC101*) [53] and its modulation of lipid A composition as a basis of conferring trifluoperazine (TFP) resistance, an inhibitor of mammalian calmodulin, to a *E. coli lpxC101* strain [54]. Since both FabG and LpxC are essential for growth and do not share homology with any known mammalian proteins, better understanding their interactions may give novel targets for the development of antibiotics that target multidrug-resistant Gram-negative pathogens.

3.2 MATERIALS AND METHODS

3.2.1 Bacterial Strains, Plasmids and Materials

The bacterial strains used were derivatives of *E. coli* K-12 (Table 3.1) and were grown in LB medium. Medium additives (in μg/ml) were sodium ampicillin, 100; spectinomycin sulfate, 100; trifluoperazine dihydrochloride, 50; and kanamycin sulfate, 50. Oligonucleotides were purchased from Integrated DNA Technologies. PCR amplification was performed using Q5 polymerase (New England BioLabs) according to the manufacturer's specifications. DNA constructs were sequenced by ACGT, Inc. Reagents and chemicals were obtained from Sigma-Aldrich and Fisher Scientific, unless otherwise noted. New England BioLabs supplied restriction enzymes and T4 DNA ligase. pCRISPR and pCas9 were a gift from Luciano Marraffini (Addgene plasmids 42875 and 42876)

The *fabGΔ8* gene was amplified via PCR from strain SW01 using primers fabG BspH1 F and fabG Sph1 R. The product was digested with restriction enzymes BspHI and SphI and inserted into the

NcoI and SphI sites of vector pBAD24 to generate pSW01. The wild type *fabG* gene was amplified from *E. coli* MC1061 using the same primers and vector to give pSW02. Similarly, plasmids pSW03 and pSW04 were constructed to carry the *fabGΔ8* gene or the wild type *fabG*, respectively. The genes were amplified using primer sets fabG BamHI F and fabG XhoI R and the products were inserted between the BamHI and XhoI sites of vector pET28a. Plasmids pSW05 to pSW08 were constructed carrying *E. coli* genes *fabD*, *fabH*, *fabA* or *fabI*, respectively, using the same protocol.

3.2.2 Construction of *E. coli* MC1061 *fabGΔ8* using CRISPR/Cas9

A CRISPR gRNA insert targeting *fabG* was synthesized by heating single-stranded oligos, CR fabG S and CR fabG AS, together for 5 minutes at 99ºC and then slowly cooling to room temperature over 2 h. The insert was then phosphorylated using T4 polynucleotide kinase. The phosphorylated insert was then ligated to pCRISPR [55], digested with BsaI, to generate pCRISPR-fabG. MC1061 transformed with pCas9 [55] (a plasmid expressing both Cas9 and the tracrRNA) and pKD46, a plasmid expressing λ-*red* under arabinose control [33], was used to prepare electrocompetent cells. These cells were transformed with 100 ng of pCRISPR-fabG and 100 ng of oligo fabG HR and allowed to recover at 25ºC for 3 h. Transformants were selected on plates containing kanamycin and the colonies were screened for temperature-sensitivity at 42ºC. Positive colonies were verified by sequencing of PCR products to confirm the expected deletion. The final strain was cured of the plasmids by serial passages in LB medium followed by screening for loss of antibiotic resistances.

3.2.3 Structural Modeling and Sequence Alignment

The FabGΔ8 sequence was structurally modeled to the *E. coli* FabG crystal structure (PDB 1I01) using Swiss-Model automated mode [56-59]. The final image was created using VMD [60]. Sequence alignments were created using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/)

3.2.4 Expression and Purification of a His-tagged FabGΔ8 Protein

Strain SW01 carrying plasmids pSW03 and pTara-Spec was grown in 1 L of LB-kanamycin medium containing 2 mM arabinose at 25°C to an OD_{600} of 0.8 and induced with 1 mM isopropyl-β-D-thiogalactopyranoside for a further 24 h. The cells were pelleted, washed in an equal volume of lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 1 mM tris(2-carboxyethyl)phosphine HCl, 10% glycerol, pH 7) and concentrated 40-fold in the same buffer. Two passes through a French pressure cell was used to lyse the cells and the suspension was centrifuged twice at 48,400 x g) for 30 min each. The cleared supernatant was applied to a 5 ml GE His-Trap HP column pre-equilibrated with the lysis buffer. The column was washed stepwise with lysis buffer containing increasing amounts of imidazole (50 mM, 100 mM, 150 mM and 200 mM) and FabGΔ8 was eluted with buffer containing 250 mM imidazole. The protein was concentrated to a final concentration of 3 mg/ml in an Amicon 10K centrifugal filter unit and dialyzed overnight against a 50 mM sodium phosphate buffer containing 500 mM sodium chloride, 1 mM tris(2-carboxyethyl) phosphine HCl, 30% glycerol, pH 7 for subsequent storage at -80°C. Wild type *E. coli* FabG was purified in a similar manner from strain MC1061 carrying plasmids pSW04 and pCY560 grown at 37°C with 2 mM arabinose to OD_{600} of 0.8 followed by induction

with 1 mM isopropyl-β-D-thiogalactopyranoside for a further 3 h. All other fatty acid synthetic proteins were purified in a similar manner following overexpression in BL21 Star for 3 h at 37°C.

3.2.5 Size Exclusion Chromatography

Purified proteins (1 mg) were applied to a Superdex 200 (10/300) column (GE Healthcare) equilibrated with a 50 mM sodium phosphate buffer (pH 7) containing 300 mM sodium chloride, at a flow-rate of 0.5 ml/min. Protein elution was monitored by absorbance at 280 nm. The mixture of molecular weight standards and Blue Dextran were run in the same manner except that using 4 mg of the mixture of standard proteins and 1 mg of Blue Dextran, respectively were loaded on the column. The void volume was calculated using the major absorbance peak of Blue-Dextran measured at 280 nm. The individual species were quantified using the area under their respective curves and expressed as a percentage of the total injected protein.

3.2.6 Isolation of Phospholipids and Lipid A

Cultures (25 ml) of strains D21, D22 and OMG3053 were grown in LB medium to an OD_{600} of 0.6 at 25˚C. The cells were harvested by centrifugation and washed with cold phosphate buffered saline. The cells were then suspended in 1ml of phosphate buffered saline and mixed with 3 ml of a 2:1 mixture of chloroform-methanol to generate a single-phase Bligh-Dyer mixture [21]. After standing for 15 min, the mixture was centrifuged at 4700 x g) for 15 min. The supernatant containing most of the phospholipids was converted to a two-phase Bligh-Dyer mixture [21] by adding 1 ml of chloroform and 1 ml of H_2O, while the insoluble material was retained to isolate lipid A. The lower phase was carefully siphoned, and the phospholipid mixture was dried under a N_2 stream and preserved at -80˚C for mass spectral analysis. The insoluble pellet containing the

lipopolysaccharide with its covalently-bound lipid A moiety, was washed once with a single- phase Bligh-Dyer mixture, consisting of $CHCl_3$/MeOH/H_2O (2:1:1 v/v/v) and again recovered by centrifugation. The washed pellets were suspended in 0.12 ml of 12.5 mM sodium acetate, 1% sodium dodecyl sulfate, pH 4.5 and sonicated briefly for 30 sec. The suspension was heated to 99˚C for 30 min, cooled and the volume adjusted to 0.36 ml. This was converted to a two-phase Bligh-Dyer mixture by adding 0.4 ml of chloroform and 0.4 ml of methanol. The phases were thoroughly mixed, and the mixture separated by centrifugation at 4700 x g) for 15 min. The lower phase was dried under a N_2 stream. To reduce the amount of SDS, the dried material was dissolved in 0.36 ml of chloroform/methanol (1:1 v/v) and subsequently converted to a two-phase Bligh-Dyer mixture by the addition of 0.4 ml of chloroform and 0.4 ml of methanol. The lower phase was recovered and dried under a N_2 stream and the resulting crude lipid A was preserved at -80˚C for further analysis by mass spectral analyses.

3.2.7 Mass Spectrometry

Electrospray ionization mass spectrometry was carried out on a Waters Quattro Ultima using the negative ion mode. Phospholipid samples were dissolved in chloroform/methanol (2:1 v/v) and infused into the source at a flow rate of 10 µl/min using a syringe pump. MS spectra were acquired with a cone voltage of 30V or 225V, over the range m/z 100-1000, in 1s. Scans over a 2 min acquisition were combined.

Matrix-assisted laser desorption ionization/time of flight mass spectra was acquired in a Brucker Daltonics UltrafleXtreme in the negative ion linear mode. The samples were prepared for MALDI/TOF by depositing 0.3µl of the lipid A sample dissolved in chloroform/methanol (2:1

v/v) followed by 0.3 µl of a saturated solution of the matrix, 2,5-dihydroxybenzoic acid in 50% acetonitrile. The sample was left to dry at room temperature. Spectra acquired were the average of 50 laser shots.

3.2.8 Radioactive Labeling of Phospholipids

The cultures were grown at 25°C in LB medium (1% tryptone, 0.5% yeast extract, 1% NaCl) to mid log phase and then shifted to 42°C. After 2 h, 5 µCi of sodium [1-^{14}C]acetate was added, and growth was allowed to continue for 2 h. The phospholipids were extracted by the method of Bligh and Dyer [21], separated by thin-layer chromatography and detected by PhosphorScreen autoradiography using a GE Typhoon FLA7000 phosphorimager. The intensity of the spots was quantified using the Image Quant TL software (GE Healthcare) relative to the wildtype. Radiolabeled fatty acids were purchased from Moravek Biochemicals. All thin-layer plates were scored into lanes to prevent cross-contamination.

3.2.9 Spectrophotometric Assay of β-Ketoacyl-ACP Reductase Activity

To measure activity spectrophotometrically, the oxidation of the cofactor NADPH was monitored at 340 nm. Reaction contained 300 µM acetoacetyl-CoA and 150 µM NADP(H) in a 100 mM sodium phosphate buffer at pH 7.0. Reaction was blanked without enzyme and FabG (0.1 µg) or FabGΔ8 (0.1 µg or 1 µg) was added. To test the effect of TFP on FabG, 10X molar excess of TFP was pre-incubated with the enzyme on ice for 15 min before addition to the reaction. The final reaction volume with enzyme was 100 µl.

3.2.10 *In vitro* Fatty Acid Synthesis Assay

β-Ketoacyl-ACP reductase activity was characterized using purified *E. coli* enzymes to catalyze steps in fatty acid biosynthesis. To produce malonyl-ACP, the fatty acid synthesis assay mixtures, which contained 0.1 M sodium phosphate (pH 7.0), 5 mM dithiothreitol, 2μg holo-ACP, and 400 μM malonyl-CoA and *E. coli* FabD (3 μg) in a final volume of 50 μL were incubated at 37°C for 1 h. Then, purified His-tagged *E. coli* proteins FabGΔ8 (2 μg/assay), and the following *E. coli* proteins at 1μg/assay FabH, FabA and FabI plus acetyl-CoA (400 μM). NADH (200 μM) and NADPH (200 μM) were added to 30 μl of the previous reaction mixture and the volume was adjusted to 40 μl with H_2O. Wild type *E. coli* FabG was used as a control. For calcium sensitivity experiments $CaCl_2$ was added to a final concentration of (0.1, 1 or 10 mM) and pre-incubated with the enzyme for 15 min at 25°C. The reaction mixtures were incubated at 25°C for 2 h (or 42°C for 45 min). The reactions were stopped by placing on ice and mixed with gel loading buffer and analyzed by conformation-sensitive gel electrophoresis on 20% polyacrylamide gels containing a 0.5 M urea [28,35]. The gels were stained and visualized using Coomassie Brilliant Blue.

3.2.11 Chemical Cross-linking of FabG

Purified mutant FabG (30 μM) was incubated at 25°C (or 42°C) for 15 min. Various concentrations of ethylene glycol bis[succinimidylsuccinate] (EGS) (0, 0.5, 2.5, 5 mM) were added and incubated at 25°C for 30 min or 42°C for 5 min. SDS-loading dye was added and samples were heated to 99°C for 5 min. Samples were loaded on 4–20% gradient SDS polyacrylamide gels (Bio-Rad) and run at 200 V for 30 min.

3.3 EXPERIMENTAL RESULTS

3.3.1 The *tfpA1* Mutation of Strain OMG3053 is a *fabG* Mutation

To identify the effects of calcium agonists on *E. coli*, Holland and coworkers. [54] isolated a TFP resistant mutant strain (OMG3053) that showed reduced growth at 42°C. This mutation, *tfpA1*, was reported to map in the *fabD* gene by complementation analysis using a genome fragment encoding the *fabD*, *fabG* and *acpP* genes of the *E. coli* FAS gene cluster. However, since no details were available as to the exact nature of the mutation, we amplified and sequenced a 2500 bp genomic region of strain OMG3053 (Primers FA F1 through FA R5), extending from within *fabH* to *fabF*. This allowed us to identify the *tfpA1* mutation as a 24 bp in-frame deletion (Δ598-621) within the *fabG* gene (Fig. 3.1A) with no changes to the rest of the FAS operon including *fabD*. The eight amino acid residues, ILAQVPAG (residues 200-207), are part of the α7 helix and the loop connecting α7 to helix α8.

Strain OMG3053 was isolated from an ill-defined parent strain (D22) having a mutation (*envA1* now *lpxC101*) that makes the cells hyperpermeable to both hydrophilic and lipophilic agents including TFP [53]. In order to study the *tfpA1* mutation in isolation, we reconstructed this deletion in the wild type *E. coli* strain MC1061 using a CRISPR-Cas9 system to produce strain SW01and renamed the mutation as *fabG*Δ*8.*

3.3.2 Phenotype of the *fabG*Δ*8* Mutation

The growth rate of the *fabG*Δ*8* mutant strain SW01in LB medium was decreased even at the permissive temperature of 25°C, with the doubling time increased to 90 min versus 60 min for the

parental strain. Growth was drastically decreased at temperatures above 30°C. However, *fabGΔ8* continues to grow for several hours, albeit very poorly, at 42°C suggesting that the cell continues to be viable even at this temperature (Fig. 3.1B). However, when grown overnight on solid media, minute colonies of the *fabGΔ8* strain form at 42°C. When these plates were returned to 25°C, growth was restored, suggesting that the cells remain viable at the non-permissive temperature. Overexpression of the mutant protein (using plasmid pSW01) in the mutant strain gave only a minor improvement in doubling time (data not shown).

The *fabGΔ8* mutation was readily complemented for restoration of growth at 42°C by a plasmid-borne wild type copy of *fabG* (pSW02). Structural modeling of the mutant protein against the wild type *E. coli* FabG indicated a severe disordering of the α6-α7 subdomain with a potential loss of interaction at a cleft formed by the end of ß6' and α5' (primes indicating neighboring monomers in the tetramer) (Fig. 3.1C). Surprisingly, *fabGΔ8* (pSW01) complemented the temperature-sensitive phenotype of CL104 (Fig. 3.2A), another *E. coli fabG*(Ts) strain that has a point mutation elsewhere in the gene. A possible explanation is that mixed dimers are formed between the two mutant proteins that compensate for one another resulting in proteins that are sufficiently stable to support growth at 42°C. Since, overexpression of *fabGΔ8* from a plasmid (pSW01) in MG1655 has no observable phenotype at 25°C or 42°C, the mutation is recessive (Fig. 3.2B) suggesting that if mixed dimers are formed between the mutant and wild type monomers, these proteins are functional.

At 25°C, unlike its mucoid *lpxC* parent, the colony morphology of strains OMG3053 and SW01 on LB agar were normal. As previously reported, the Ts mutation of strain OMG3053 confers

resistance to 50 μg/ml TFP in the D22 background (Fig. 3.2C). However, upon prolonged incubation strain OMG3053 colonies become mucoid like the parental strain D22 whereas strain SW01 retains normal colony morphology. Unexpectedly, strain SW01 gave revertants that grew reliably at 42°C. Revertants were isolated at a frequency of about 10^{-8}. The fatty acid operons of sixteen revertants were sequenced with twelve identified as intragenic suppressors in *fabG* (M96V or E185D). The eight-residue deletion remained unaltered. Met96 is located at the end of helix α4 at the α4-α4' dimer interface whereas Glu185 is located at the mouth of the active-site tunnel in proximity to α6. The other suppressors are in locations other than the fatty acid operon and are yet to be identified. One possibility is that these suppressors increase the levels of chaperone proteins which can stabilize mutant proteins [61,62].

3.3.3 *In vitro* β-ketoacyl-ACP Reductase Activity/Calcium Sensitivity

The reductase activity of FabG is readily detected spectrophotometrically by monitoring the oxidation of the NADPH cofactor at 340 nm [43]. Although wild type FabG can reduce acetoacetyl-CoA using NADPH, FabGΔ8 showed no activity towards CoA-thioester substrates (Fig. 3A). Note that the FabG acetoacetyl-CoA reductase activity is not inhibited by TFP indicating that the action of TFP is indirect. To characterize the activity of FabGΔ8 *in vitro* using its physiologically-relevant acyl-ACP substrates, a reconstituted fatty acid synthesis system was utilized (Fig. 3.3B). Due to its instability, malonyl-ACP was generated in the reaction using purified *E. coli* FabD and then converted to acetoacetyl-ACP using *E. coli* FabH and acetyl-CoA. For product stability and to obtain sufficient separation on conformation-sensitive gels, the product of the FabG reduction was further dehydrated and reduced to butyryl-ACP using *E. coli* FabA and FabI. At 25°C, both

FabGΔ8 and FabG can catalyze reduction of the acetoacetyl-ACP, whereas at 42°C, FabGΔ8 is inactive. (Fig. 3.3C)

Price and coworkers[63] have shown that *E. coli* FabG is sensitive to the presence of divalent cations of calcium and magnesium, with monovalent cations of sodium having no effect on activity. Their structural data seems to suggest that these cations might block *in vitro* activity by occluding access to active site. One of the calcium binding sites in the tetramer is formed between the Gln203 located in the α6/α7 subdomain and the oxygen of Asn145 [63]. The lack of the α6/α7 subdomain in FabGΔ8 which includes Gln203 results in a 100-fold increase in the sensitivity of FabGΔ8 to the presence of calcium at 25°C. The β-ketoacyl-ACP reductase activity of FabGΔ8 was significantly inhibited by100 μM calcium whereas inhibition of the wild type FabG required 10 mM calcium under the same conditions. Monovalent sodium cations had no observable effect (data not shown) as expected from prior work [63].

3.3.4 The FabGΔ8 protein is Deficient in Dimer Formation at 42°C

The absence of part of the α6/α7 subdomain suggested that the dimerization ability of the mutant protein should be compromised. A C-terminal hexahistidine-tagged version of FabGΔ8 protein was overexpressed in the *fabGΔ8* mutant background (SW01) strain at 25°C using a helper plasmid that produced T7 RNA polymerase. The hexahistidine-tagged wild type *E. coli* FabG protein is readily over-expressed, in soluble form and readily purified using routine Ni-NTA based methods[63]. The FabGΔ8 protein, like the proteins encoded by other *fabG*(Ts*)* strains[49] was difficult to purify in an active form and was very labile. Indeed, earlier work on other *fabG*(Ts*)* strains was limited to assaying enzyme activities in cell free extracts due to the extreme lability of the mutant

enzymes. We purified the FabGΔ8 to homogeneity using Ni-NTA affinity chromatography and the protein was immediately dialyzed to remove the eluting imidazole, concentrated 10-fold and stored at -80°C. Handling time was minimized at the expense of protein recovery to facilitate retention of activity. The wild type FabG was purified similarly from the parental strain (MC1061). SDS-PAGE analysis showed that the mutant protein migrated with an apparent molecular weight of 25 kDa (Fig. 3.4A) whereas the wild type protein migrated with an apparent molecular weight of 27 kDa. Size exclusion chromatography of the purified FabGΔ8 protein indicated that in its solution form ~52% of the total protein was monomeric (apparent molecular mass of 23 kDa) with only ~27% present as a tetramer (92 kDa). About 21% of the protein was observed in an aggregated form. The wild type protein, however, was predominantly dimeric (~99%) with an apparent molecular mass of 48 kDa (Fig. 4B) with ~1% aggregation. This indicates that the mutant has impaired interactions that prevent it from undergoing the first dimerization event whereas the second dimerization event that forms the tetramer proceeds well. It seems likely that in the small fraction of the mutant protein that manage to form the first dimer, the second dimerization event has a compensating effect that stabilizes the tetrameric structure. To test the effects of temperature, the purified proteins were incubated at 25°C and 42°C for 30 min and then cross-linked using varying amounts of a chemical crosslinking agent. The crosslinking reactions were quenched and analyzed by SDS-PAGE. Although the mutant protein, like the wild type FabG protein, showed dimerization and tetramerization at 25°C, the mutant required a 10-fold higher concentration of crosslinker to produce a visible a dimer band, indicating a significant dimerization deficiency at 42°C, although tetramer formation remained unimpaired at this temperature. Of note, the wild type protein has a more prominent dimer band than its tetramer band even at higher concentrations of the crosslinker.

These data, in combination with structural information about the monomer-monomer interaction interfaces, offers a glimpse in the mechanism of FabG tetramerization. These data suggest that in the intact protein the most stable and probably the first interaction occurs between the ß7 strands with some interactions involving the α6/α7 subdomain. The second dimerization at the α4/α4' and α5/α5' dimer interface to form the tetrameric form further stabilizes the structure. The FabGΔ8 protein is impaired at the first dimerization stage and thus either remains as an inactive monomer or requires the additional stabilizing interactions given by tetramerization to form a functional protein.

3.3.5 The *fabGΔ8* strain SW01 is Defective in Phospholipid Synthesis at the Non-permissive Temperature and Shows Altered Fatty Acid Composition

In bacteria, a majority of the fatty acids produced in the cell are used for the synthesis and maintenance of the phospholipid bilayers and the lipid A component of the lipopolysaccharide. To identify the reason for the reduced growth of the *fabGΔ8* mutant at 25°C and the growth deficiency at 42°C, cultures were grown to mid log phase at the permissive temperatures and then shifted to 42°C for 2 h. ^{14}C-Acetate was added to the cultures which were growth for an additional 2 h. The cultures were normalized by optical density measurements at 600 nm and cellular phospholipids were then extracted and separated by thin layer chromatography and visualized by autoradiography. The levels of fatty acid synthesis *in vivo* after the temperature upshift could then be quantitated by measuring the radiolabel incorporated into the phospholipid fraction. The mutant retained only 30% of the fatty acid synthetic ability of the wild type strain (Fig. 3.5A). The loss of

synthesis was successfully complemented by introduction of a wild type copy of *fabG* on plasmid pSW02 into the mutant strain.

The composition of fatty acids species in the *fabG*Δ*8* mutant strain SW01 grown at 25°C was measured by in-source collision induced electrospray ionization dissociation of the phospholipid samples. The data showed that relative to the wild type strain, MC1061, the levels of the C18:1 fatty acid decreased by 60% whereas C16:0 decreased by 25-30% in the mutant relative to the major C16:1 peak in comparison to the wild type strain (Fig. 3.5B). Electrospray ionization analysis of the molecular species of phosphatidylglycerol gave similar results with a shift to species having shorter acyl chains (data not shown) as expected from reduced flux through the fatty acid synthesis cycle.

3.3.6 The *fabG*Δ*8* Mutation Alters the Lipid A Fatty Acid Composition but Only in the *lpxC101* Background

The acylated lipid A core is an essential component of the Gram-negative cell wall that anchors the lipopolysaccharide backbone to the outer membrane. Lipid A was isolated and analyzed by MALDI-MS using a gentle SDS based method [64,65]. In agreement with previous reports for wild type *E. coli*, a major peak at m/z 1797.16 corresponding to a hexa-acylated diphosphorylated lipid A (Figs. 3.6A and 6B) containing four C14:0 3-hydroxy chains and one C12:0 chain. Other minor peaks centered at m/z 1768.13 and m/z 1824.20 correspond to a loss/gain of a hydroxyl group, whereas those centered at m/z 1716.16 correspond to a loss of a phosphate group. In combination with the *lpxC101* mutation, the *fabG*Δ*8* strain (OMG3053) produced a remarkably different set of lipid A species. The negative mass spectrum was dominated by four sets of peaks centered at m/z

1445.87, m/z 1615.98, m/z 1797.16 and m/z 1852.17. The peak at m/z 1615.98 was assigned to hexa-acylated diphosphorylated lipid A species having two C10:0 3-hydroxy chains, three C12:0, 2 (or 3-OH C12:0 chains) and one C12:0 chain (Figs. 3.6A and 6D). The peaks centered at m/z 1445.87 correspond to the loss of one C10:0 3-hydroxy acyl chain (Δ_{mass} of 170) whereas those at m/z 1852.17 correspond to the gain of one C16:0 acyl chain (Δ_{mass} of 236). Strain D22 (*lpxC101*) retained a wild type lipid A composition (Fig. 3.6C). Remarkably, strain MC1061 derived *fabGΔ8* strain SW01, had a wild type lipid A acyl chain composition suggesting that a tight regulatory network exists between FAS and lipid A synthesis that maintains structural integrity of the cell, that that cannot be disrupted by inactivation of a single gene in either pathway.

3.4 DISCUSSION

In this chapter, I delineate the biochemical properties of a *E. coli fabG*(Ts) mutant strain and its phenotypic effects on growth. Unlike typical temperature-sensitive phenotypes caused by point mutations, the *fabGΔ8* strain lacks eight residues that form a part of the α7 helix and the loop connecting α7 to the α8 helix in the α6/α7 subdomain. This subdomain is responsible for active-site tunnel formation and for dimerization [47].

My data shows that FabGΔ8 cannot efficiently form dimers, probably due to weakened monomer-monomer interactions at the dimer interface that involve the α6/α7 subdomain and Tyr129' that is further exacerbated by an increase in temperature and leads to significant reduction in enzyme activity. Although *E. coli* FabG crystallized as a tetramer, to my knowledge the only measurements of the solution structure of a ß-ketoacyl-ACP reductase were done on a plant enzyme [66].

Using exclusion chromatography, I show that the protein predominantly exists as a dimer in solution. This is further validated by the more prominent dimer band in my crosslinking experiments. The mutant protein, however, exists either as a monomer or a tetramer. Since, the FabG tetramer is well described as a dimer of dimers [47], my data suggests that the initial dimerization event is the limiting step in assembly of the mutant tetramer.

Interestingly, the altered interface also appears to play an important role in NADP(H)-induced allostery[63]. In the mutant protein, the additional interactions resulting from the second dimerization (dimerization of two dimers) seem to act to stabilize the protein in the active tetrameric form. Unexpectedly, the deletion mutant reverts at a frequency reminiscent of a point mutant. One of the suppressor mutations (M96V) maps to the second dimerization interface. The interaction at this interface is mainly hydrophobic with one of the suppressors replacing a methionine side chain with a differently shaped and somewhat more hydrophobic valine side chain. M96 participates in an important hydrogen bonding interaction with Glu168' which plays a role in the allosteric behavior of FabG. This bonding probably increases the interaction between the dimers offering sufficient stabilization at higher temperatures to restore function.

The second intragenic suppressor observed (E185D) is at the base of the α6/α7 subdomain. Substituting glutamate with the shorter aspartate might serve to rein in the α6 domain by interaction with Arg190 to give increased order of the interface. Also, since the mutant strain retains the ability to grow very slowly at 42°C, a small fraction of the monomers probably interacts long enough to form active tetramers that catalyze the residual growth of the mutant strain. It should be noted that there is no *a priori* reason (such active sites formed by an interface) for FabG

to be a tetramer. Indeed, a dimeric ß-ketoacyl-ACP reductase is encoded by *Synechocystis* [67] and the solution structure of *Pseudomonas aeruginosa* RhlG, a ß-ketoacyl-ACP reductase essential for rhamnolipid synthesis, has been reported to be dimeric by both size exclusion chromatography and analytical ultracentrifugation, although at high protein concentrations tetramers were observed [68].

Although the mutant strain OMG3053 originally isolated by Holland et al. [54], was reported to a *fabD* mutation that conferred TFP resistance to an *lpxC* strain, the report was in error. This report is puzzling in that, although the investigators sequenced a subclone derived from the original complementing plasmid that carried the wild type sequence, they did not report the sequence of the mutant gene. Moreover, a plasmid carrying the entire *fabD* gene together with upstream (*fabH*) and downstream (*fabG*) sequences failed to restore growth at 42ºC although transcription should have been driven by the vector tetracycline promoter [54]. A plasmid encoding *fabG* also failed to complement, however that gene lacked a promoter. The *fabG* gene was encoded on the strand opposite the tetracycline promoter and later work showed that the cloned fragment does not contain a promoter [48]. Indeed, the authors concluded that it "has not been unequivocally demonstrated that *tfpA1* mutant carries a single mutation in *fabD*".

LpxC catalyzes an irreversible deacetylation step during lipid A synthesis. The *lpxC101* mutation results in sensitivity to TFP in addition to other agents like detergents. Moreover, strain D22 has a very abnormal, mucoid and slimy colony morphology. Since *lpxC101* mutants show marked outer membrane defects, we constructed the *fabG*Δ*8* mutation in a wild type strain using a markerless CRISPR-Cas9 based methodology. Although both *lpxC101*(D22) and *fabG*Δ*8* (SW01) have growth defects, they have normal lipid A compositions. Remarkably, strain OMG3053 has a very

unusual lipid A composition that has C10 and C12 chains in place of C14 chains to give composition that is reminiscent of *Pseudomonas aeruginosa* lipid A [69]. It is known that *E. coli* can tolerate a similar shortening of the lipid A acyl chains [70]. *In vitro*, TFP has no effect on the activity of FabG measured spectrophotometrically using acetoacetyl-CoA as a substrate, suggesting that TFP does not directly affect FabG. Earlier studies have identified *E. coli* suppressors in *fabZ* that compensate for the presence of LpxC inhibitors and restore cellular homeostasis [71]. This has been attributed to the fact that since both *fabZ* and *lpxD* (which produces the LpxC substrate) compete for the same ß-hydroxymyristoyl-ACP substrate pool, a compensating mutation in *fabZ* would lead to a balanced growth in a *lpxC* mutant. However, suppressor mutations in *E. coli fabG* were not reported. Surprisingly, in *P. aeruginosa* similar LpxC inhibitor studies gave, mutations in *fabG* [72]. It should be noted that FabG and FabZ catalyze consecutive steps in the fatty acid synthesis cycle and both reactions are freely reversible. Hence, interplay between the two enzyme activities does not seem surprising.

3.5 TABLES AND FIGURES

Strains	Relevant Genotype	Source
MC1061	Wild type strain	Casadaban, et al., 1980
DH5α	*endA1 recA1 0 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17*	Hanahan, et al., 1985
MG1655	*K-12 F– λ– ilvG– rfb-50 rph-1*	CGSC #6300
BL21 (Star)	*ompT hsdSB (rB-, mB-) gal dcm rne131 (DE3)*	Invitrogen
D21	*proA23, lac-28, tsx-81, trp-30, his-51, rpsL173 ampCp-1*	Boman, et al., 1968
D22	As D21, *lpxC101*	Normark, et al., 1969
OMG3053	As D22, *tfpA1*	Bouquin, et al., 1995
CL104	MG1655 *fabG*(Ts) *panD::Cm*	Lai, et al., 2004
SW01	MC1061 *fabGΔ8* (Ts)	This work

Table 3.1 Bacterial Strains used in this study

Plasmids	Relevant Characteristics	Source
pBAD24	Amp^R, Arabinose-inducible expression vector	Guzman, et al., 1995
pET28a	Kan^R, T7 promoter-based, IPTG inducible expression vector	Novagen
pCY560	Spc^R, Arabinose-inducible expression of T7 Polymerase;	Hassan, et al., 2011
pSW01	Amp^R, PCR-amplified *fabG*Δ*8*(Ts) gene from SW01 in Nco1/SphI sites of pBAD24	This work
pSW02	Amp^R, PCR-amplified wild-type *fabG* gene from MC1061 in NcoI/SphI sites of pBAD24	This work
pSW03	Kan^R, PCR-amplified *fabG*Δ*8*(Ts) gene from SW01 in BamHI/XhoI sites of pET28a	This work
pSW04	Kan^R, PCR-amplified wild-type *fabG* gene from MC1061 in BamHI/XhoI sites of pET28a	This work
pSW05	Kan^R, PCR-amplified *fabD* gene from MG1655 in BamHI/XhoI sites of pET28a	This work
pSW06	Kan^R, PCR-amplified *fabH* gene from MG1655 in BamHI/XhoI sites of pET28a	This work
pSW07	Kan^R, PCR-amplified *fabA* gene from MG1655 in BamHI/XhoI sites of pET28a	This work
pSW08	Kan^R, PCR-amplified *fabI* gene from MG1655 in BamHI/XhoI sites of pET28a	This work
pCRISPR	Kan^R, crRNA expression plasmid	Jiang et. al., 2013
pCRISPR	Kan^R, fabD crRNA inserted in Bsa1 sites	This work
pCas9	Cm^R, *Streptococcus pyogenes* Cas9 nuclease and tracrRNA under their native promoters	Jiang et. al., 2013
pKD46	Amp^R, Ts ori, Arabinose-inducible expression of λ-Red genes	Datsenko and Wanner, 2000

Table 3.2 Plasmids used in this study

Oligonucleotides	Sequence
fabG BspH1 F	ATGCATTCATGAATTTTGAAGGAAAAATCGC
fabG Sph1 R	ATGCATGCATGCTCAGACCATGTACATCCCGC
fabG BamH1 F	ATGCAT GGATCC ATGAATTTTGAAGGAAAAATCGC
fabG Xho1 R	ATGCATCTCGAGTCAGACCATGTACATCCCGC
FA F1	GATCGCCACGGTAATACCTC
FA F2	TGCGTCTGGTTGAGATGC
FA F3	GCTTACTGGCCTGACGAAAC
FA F4	CACTATCGGTTCTGTGGTTGG
FA F5	CTACGAAAACCATCGCGAAA
FA R1	CTTTCGCAATAGACGCATCA
FA R2	CTTTTAAAGCTCGAGCGCC
FA R3	TTACTGAAGCCGATCAAGCC
FA R4	CAGCTGTTCGCCGATAATTT
FA R5	AATGAAGGCATCCATCTTGC
CR fabG S	AAAC TATCCTGGCGCAGGTTCCTG G
CR fabG AS	AAAAC CAGGAACCTGCGCCAGGATA
fabG HR	ACCGCGTTGGCGATTTCCTGTGCGCCGCCGAGGCGACCCGCAC GCTGGTCATCGCTCAGCGCACGTGTCA

Table 3.3 Oligonucleotides used in this study

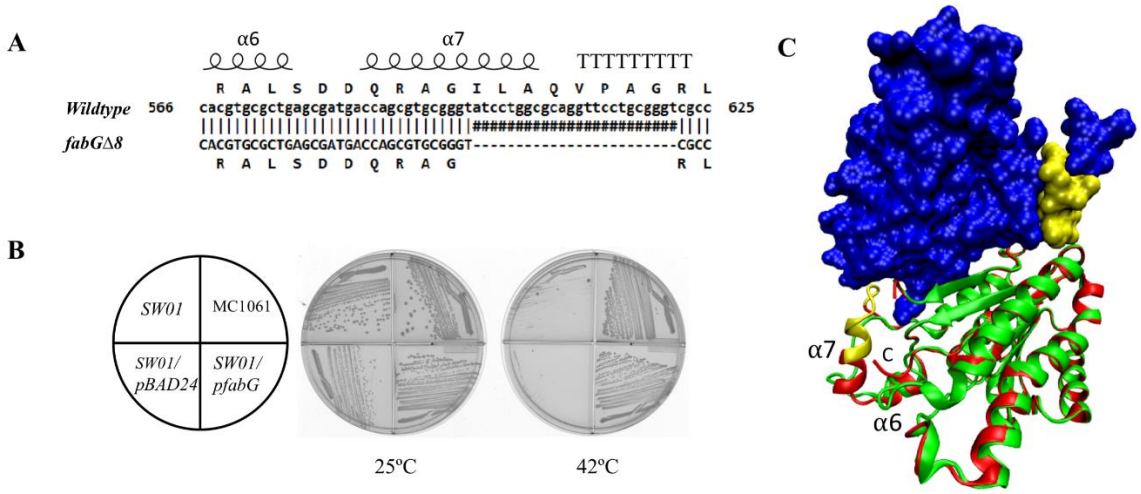

Figure 3.1 The *E. coli fabGΔ8* encoded protein lacks eight amino acid resides and confers temperature sensitivity and trifluoperazine resistance. A) Sequence alignment of bases 566 to 625 of *fabG* from strain MC1061 and the *fabGΔ8* strain SW01. Missing bases are indicated by '#'. Helices α6 and α7 and the turn (indicated by a series of 'T's) are shown above the amino acid sequence **B)** Growth of mutant *E. coli* strain *fabGΔ8* (SW01*)* at 30ºC or 42ºC. The control strains were the wild type MC1061, strain SW01 with either an empty pBAD24 vector or this vector bearing the wild type *E. coli fabG* (pSW02) **C)** Structural model of FabGΔ8 (green) generated by SWISS-MODEL aligned with a monomer of wild type FabG (red) (PDB:1I01). A space-filling model of a second monomer is shown to highlight monomer-monomer interactions missing in FabGΔ8. Missing residues are shown in yellow in the wild type subunits.

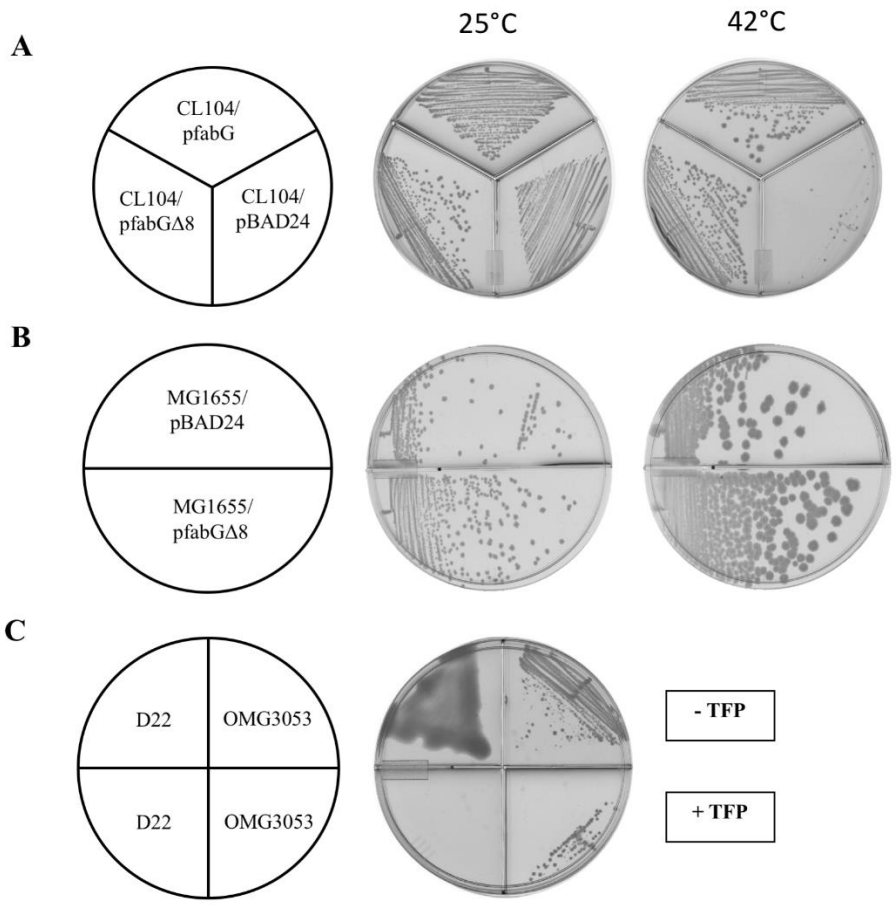

Figure 3.2 Expression of *fabGΔ8* complements growth of the *E. coli fabG*(Ts) strain CL104 at 42ºC, *fabGΔ8* is recessive to wild type *fabG* and causes resistance to TFP in an *lpxC101* background. A) CL104 is an *E. coli fabG*(Ts) strain carrying a E233K point mutation[49] . FabGΔ8 was expressed from a pBAD24-derived plasmid from an arabinose inducible promoter. The control strains are CL104 containing the empty vector (pBAD24) or pBAD24 expressing wild type *E. coli fabG* (pSW02). Complementation proceeded both in the presence and absence of arabinose. **B)** Expression of the mutant FabG (pSW01) does not effect growth of the wild type *E. coli* strain MG1655 at 25ºC or 42ºC. FabGΔ8 was expressed from a pBAD24-derived plasmid under an arabinose inducible promoter. The control strain is strain MG1655 containing an empty vector (pBAD24). Growth was observed both in the presence and absence of arabinose. All plates contained arabinose. **C)** Growth of mutant *E. coli* strain OMG3053 in the presence and absence of 50 µg/ml TFP. Strain D22 served as a control.

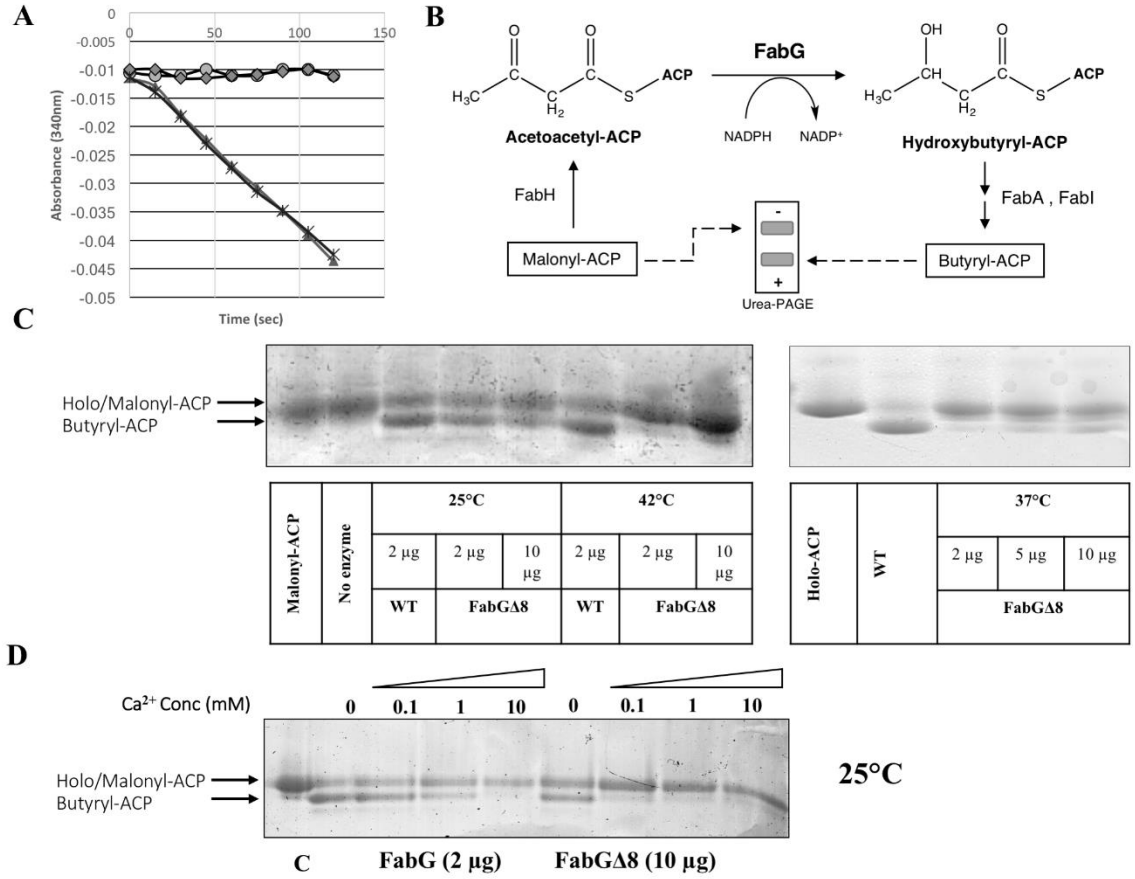

Figure 3.3 FabGΔ8 loses ß-ketoacyl-ACP reductase activity at higher temperatures and in the presence of calcium. A) Oxidation of the NADPH cofactor was monitored spectrophotometrically at 340 nm for 120 seconds. Symbols ▲: FabG (0.1 µg) ●: FabGΔ8 (0.1 µg) ◆: FabGΔ8 (1 µg) ×: FabG + TFP **B)** Schematic diagram of the enzymatic reaction system used to test ß-ketoacyl-ACP reductase activity. **C)** Enzymatic assays for FabGΔ8 catalyzed reduction of acetoacetyl-ACP to hydroxyl butyryl-ACP. The enzymatic reactions (40 µl) were performed at 25°C, 37°C or 42°C and contained holo-ACP converted to acetoacetyl-ACP by the action of *E. coli* FabD (plus malonyl-CoA) and FabH (plus acetyl-CoA). The product is dehydrated (FabA) and then reduced (FabI) to the faster-migrating butyryl-ACP, which can be resolved from the substrate in a destabilizing urea-PAGE system (Materials and Methods). FabGΔ8 was tested at two levels (2 µg and 10 µg) at 25°C and 42°C and at 3 levels (2 µg, 5 µg and 10 µg) at 37°C.

Figure 3.3 (cont.)

The no enzyme controls contained all components except FabG or FabGΔ8. **D)** Effects of calcium on FabGΔ8 reductase activity *in vitro*. 'C' denotes no addition of FabG/FabGΔ8 whereas the triangle represents an $CaCl_2$ dilution series (0, 0.1, 1, 10 mM). Reactions are performed as above. Lanes 2-5 had 2 μg of FabG added per reaction (40 μl) whereas lanes 6-9 had 10 μg of FabGΔ8 added per reaction.

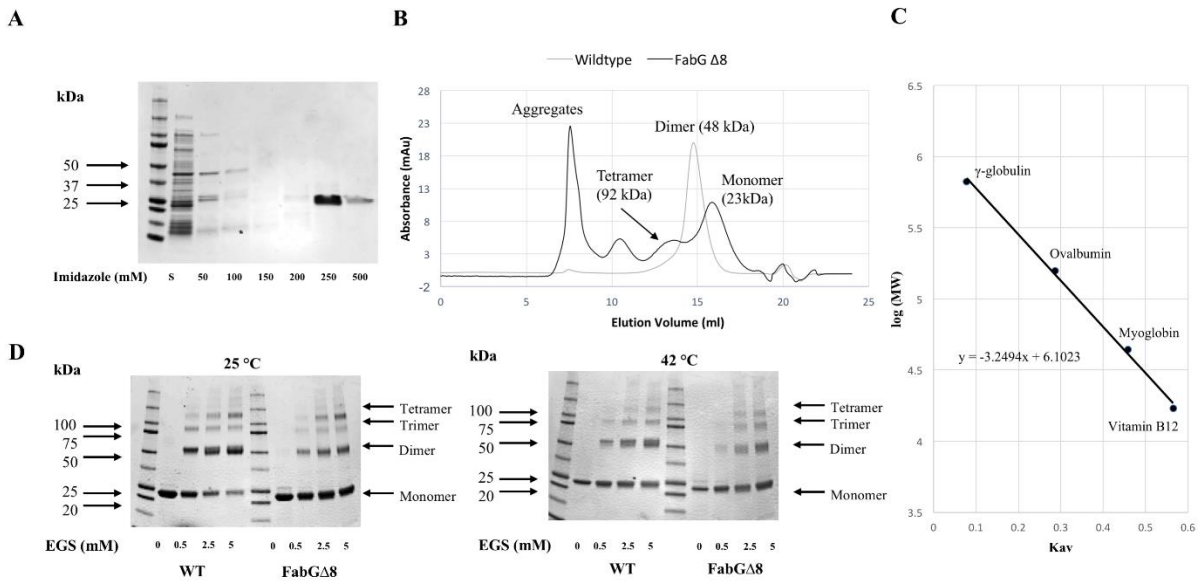

Figure 3.4 FabGΔ8 is deficient in dimerization. A) Purification of C-terminally His_6-tagged FabGΔ8 protein and analysis by SDS-PAGE. Lane 2: Extract of the FabGΔ8 overexpression strain; Lanes 3-8: Serial elution with 50-500 mM Imidazole; Lane 7: Purified FabGΔ8 protein. **B)** Size exclusion chromatographic elution profile of purified hexahistidine-tagged FabGΔ8 (in black). This is overlaid with a similar profile from purified hexahistidine-tagged wild type FabG (in grey), with the estimated molecular masses and their solution structure indicated next to each peak. **C)** The solution structures of the proteins were determined by comparison to the elution patterns of a series of standards (Bio-Rad) and Blue Dextran. The standards were vitamin B12 (1.35 kDa), myoglobin (horse, 17 kDa), ovalbumin (chicken, 44 kDa), γ-globulin (bovine, 158 kDa). The elution position of each of the standard peaks gave an estimated molecular mass marked on the profile based on graphic analysis of the standard curve. Kav, partition coefficient. MW, molecular weight. **D)** Chemical crosslinking of FabG and FabGΔ8 proteins with increasing amounts of ethylene glycol bis[succinimidylsuccinate] (EGS) (0 to 5 mM). **Left panel:** Incubated at 25°C for 15 min followed by crosslinking for 30 min at 25°C **Right Panel:** Incubated at 42°C for 15 min followed by crosslinking for 5 min at 42°C.

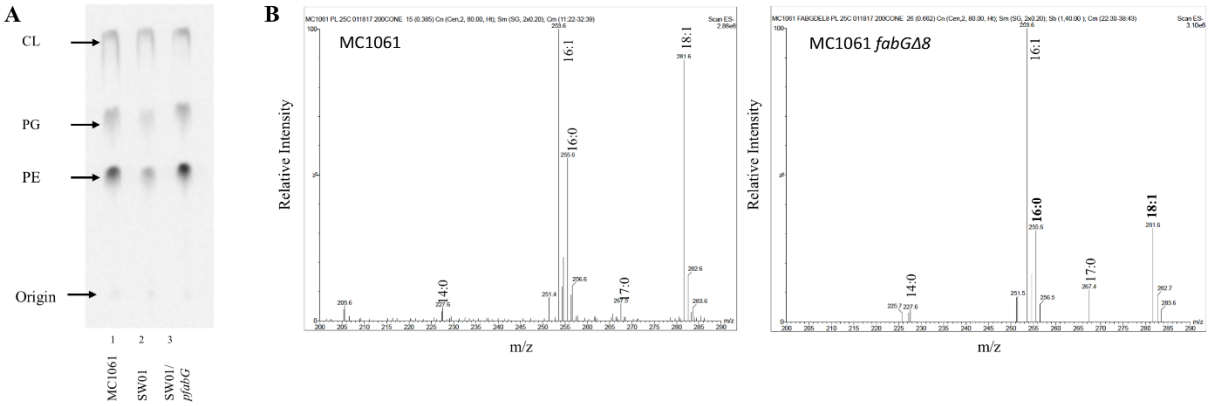

Figure 3.5 Changes in the synthesis and composition of phospholipid fatty acids synthesized by the *fabG*Δ*8* SW01strain. A) Autoradiograph of radiolabeled phospholipids separated on a silica G-thin layer chromatography plate. Phospholipids were labelled using [^{14}C]acetate added to the growth medium after shift from 25°C to 42°C. Strains MC1061 and SW01 complemented with wild type *fabG* (pSW02) served as controls The amounts of phospholipids produced relative to the wild type MC1061strain at 42°C were 30% for strain SW01 and 116% for strain SW01 carrying the FabG wild type expression plasmid pSW02. The radiolabeled thin layer chromatography spots were quantified with a GE Typhoon FLA7000 phosphorimager using the Image Quant TL software and all values are the average of two experiments. **B)** Electrospray ionization mass spectrometry of phospholipid extracts of the wild type strain MC1061 (left panel) and the *fabG*Δ*8* SW01 strain (right panel) grown at 25ºC. Fatty acids with >10% change are given in bold.

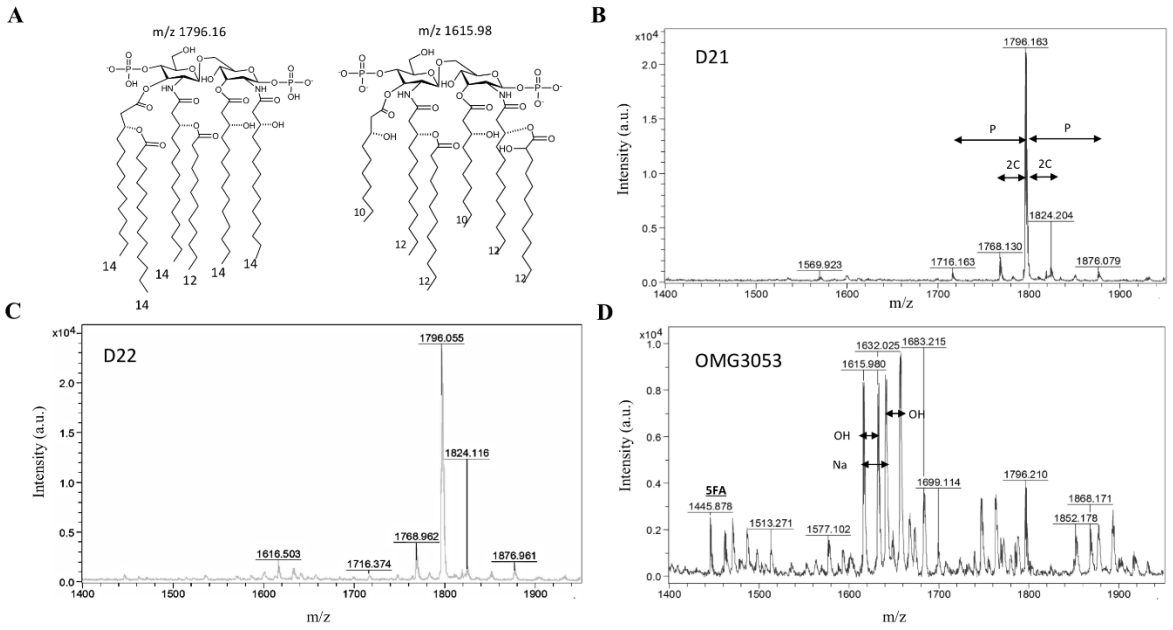

Figure 3.6 Lipid A composition of the mutant in an *lpxC* background. A) Predicted structures of the major lipid A species observed, with chain length indicated below each acyl chain. **B)** Negative ion mode MALDI-TOF spectra of lipid A extracts from D21 (wildtype) **C)** D22 (*lpxC101*) **D)** OMG3053 (*lpxC101 fabG*Δ*8*). (**P**: phosphoryl group; **OH**: hydroxyl group; **Na**: sodium adducts; **2C**: plus two carbons; **5FA**: penta-acylated species)

Chapter Four

A RAPIDLY CURABLE CRISPR/CAS9 SET OF VECTORS BASED ON AN IPTG-DEPENDENT REPLICON

4.1 INTRODUCTION

The ability to make sophisticated and precise genome modifications in a reasonable timescale has made *Escherichia coli* one of the most comprehensively studied and understood genomes of any organism. Most of the first available genome engineering tools in *E. coli* rely on either a phage-derived homologous recombination (HR) pathway [4-6] or group-II intron-based homing [7] for targeted gene editing. Recombinogenic methods based on the λ-phage proteins Exo, Beta and Gamma (λ Red)[8] catalyze allelic exchange with just 35 bp of homology but typically utilize an antibiotic selection marker and leave a "scar" behind . Achieving scar-less manipulation usually requires a counter-selectable marker or multiple rounds of recombination. Despite the successful application of these techniques, modification of genes that are either themselves essential or are part of an operon with downstream essential genes remains challenging. Most of the current tools either require expression of the gene(s) in *trans* or a more complex cloning strategy to restore disrupted codons/promoters.

One of the most recent breakthroughs has been the development of a novel RNA-guided endonuclease-based genome editing system called the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR associated *Streptocccus pyogenes* (Sp) protein 9 system (Cas9). A CRISPR RNA (crRNA) forms an RNA duplex with a trans-acting crRNA (tracrRNA),

which then helps recruit the Cas9 endonuclease to target a 20 bp protospacer on the genome which matches the sequence of the crRNA and is immediately adjacent to a Cas9 specific protospacer adjacent motif (PAM). Later studies demonstrated that the tracrRNA:crRNA duplex can be replaced with a single guide RNA (sgRNA). Thus, by simply changing the 20 bp sequence within a plasmid-encoded sgRNA, one can reprogram the target of the Cas9 endonuclease. The DSBs resulting from CRISPR action are lethal to *E. coli* because it lacks a non-homologous end joining mechanism (NHEJ).

The CRISPR/Cas9 system using crRNA expressed from a plasmid has been shown to successfully incorporate point mutations in the *rpsL* gene in *E. coli* in the presence of λ-Red with an efficiency of 60%. Early varieties of the CRISPR/Cas9 tools in *E. coli* provided SpCas9 and its tracrRNA under its native promoter on one plasmid and a crRNA in a separate high copy number plasmid, with a requirement for the presence of the λ*red* genes on a third plasmid or on the chromosome. Later iterations have used sgRNA for ease of cloning and combined λ -Red with either the Cas9 gene or the sgRNA.

Curing these plasmids after accomplishing the desired changes often requires multiple rounds of growth in selection-free media followed by a screening for plasmid loss. Other two-plasmid systems typically utilize a pSC101(Ts) backbone for the sgRNA plasmid to permit easier curing. However, these plasmids cannot be used to generate the Ts mutations utilized to study essential genes.

In this Chapter, I describe the construction of a set of compatible vectors which can only replicate in the presence of IPTG. Using one of these vectors, a CRISPR-Cas9 system was assembled that can be used to make rapid iterative changes with ease while leaving a plasmid-free modified strain.

4.2 MATERIALS AND METHODS

4.2.1 Bacterial Strains, Plasmids and Growth Conditions

E. coli strain BW25113 was employed for genome editing and DH5α for cloning. Lambda Red recombineering donor plasmid pKD46 was obtained from the *E. coli* Genetic Stock Center at Yale University. pCas9-CR4 and pKDsgRNA-p15a were a gift from Kristala Prather (Addgene plasmids # 62655 and 62656). Oligonucleotides were synthesized by Integrated DNA Technologies (IDT) at the 25 μg scale using standard desalting. *E. coli* strains were grown aerobically in shake flask cultures at 37˚C and 250 rpm in lysogeny broth (LB) unless stated otherwise. Recombinant strains were selected with 100 μg/ml ampicillin or 10 μg/ml chloramphenicol and stored in 40% glycerol at -80˚C.

4.2.2 Construction of the pI Vector Series

The p15a origin was amplified from vector pACYC184 by PCR using primers p15a KpnI F and p15a XbaI R while the RSF1030 and RSF1031 origins were amplified by PCR from pCY2016 and pCY2017 using primers RSF KpnI F and RSF XbaI R. RSF1030 and RSF1031 differ from one other by a single nucleotide. The primers contain restriction sites for KpnI and XbaI, respectively. The respective fragments and pAM34 were digested with KpnI plus XbaI and ligated to obtain plasmids pI15, pI1030 and pI1031, respectively. Antibiotic resistance markers for

chloramphenicol, kanamycin and gentamycin were amplified from p34-C, p34-K and p34-G by PCR using the primers p34C AatII F/ p34C Xba1 R, p34K AatII F/ p34K XbaI R and p34G AatII F/ p34G XbaI R, respectively. The resultant fragments were digested with AatII/Xba1 and ligated to a pAM34 fragment digested with the same restriction enzymes. For consistency, the resultant plasmids were named pIBRC, pIBRK and pIBRG.

4.2.3 Determination of Copy Numbers and Rates of Curing

Copy number determinations were made by growing cultures in LB to an OD of 1.0 and extracting plasmids from 1 ml of culture. The extracted plasmids were digested with XbaI and run on a 1% agarose gel. DNA amounts were estimated by densitometry calculations on a Biorad Chemidoc XRS.

To measure the rate of curing of plasmids pAM34, pI15, pI1030 and pI1031, strains carrying the plasmids were first grown in a selection-free LB medium and then periodically plated on LB agar containing IPTG. Optical density measurements were also made at this time to calculate the number of generations that occurred since selection pressure was removed. Cultures were again diluted when an OD of 1.2 was reached until 25 generations had elapsed. Thirty six colonies were screened per plasmid, at each timepoint, for resistance to ampicillin.

4.2.4 Construction of pITRCas9 and pSgRNA

A fragment containing *lacI*^{Q} and part of the p15a origin was amplified by PCR from pI15 using primer p15a-lacI SpeI F and p15a-lacI AgeI R. pCas9-CR4 and the PCR fragment were digested with SpeI plus AgeI and ligated to give pSW09. λ*red* genes under arabinose control were amplified

from pKD46 using primers LRed AatII F and LRed term AatII R. A minimal dual direction tonB terminator was introduced with sequence 5'-CATGGTAATAGTCAAAAGCCTCCGGTCGGAG GCTTTTGACTTAAGCTT-3'. This fragment was digested with AatII and ligated with pSW09 digested with the same enzyme. The resultant plasmid was named pITRCas9. Correct assembly was confirmed by restriction digestion and by PCR amplification of the junctions.

To construct pSgRNA, pKDSgRNA-p15a and an ampicillin resistance fragment amplified from pKD46 (primers Amp NcoI F and Amp XhoI R) were digested with NcoI plus XhoI and ligated to give pSW10. This removed the plasmid encoding the λ*red* genes and spectinomycin resistance fragment. The resulting plasmid was linearized by PCR using primers Non Ts F and Non Ts R while introducing a point mutation (A to G) that restores temperature independent growth. The linear fragment was phosphorylated using NEB T4 polynucleotide kinase and self-ligated to give pSW11. This was digested with NcoI and ligated with a synthetic construct containing an *E. coli* codon optimized *Photorhabdus luminescens rpsL* gene under a constitutive *proD* promoter. The 20 bp protospacer region of the resulting plasmid pSgRNA-p15a, was modified to contain a NcoI site flanked by two BbsI sites by PCR using the strategy described below to yield pSgRNA.

4.2.5 Protospacer Cloning into pSgRNA

All pSgRNA-xxx plasmids (where xxx denotes the gene targeted) were created using a one-step digestion-free PCR based cloning (DFPC) strategy as described below. The entire pSgRNA plasmid was amplified via PCR using a target specific primer pSgRNA-xxx F, a fixed primer pSgRNA-R and pSgRNA as template to yield a linear product. A 15 cycle PCR was performed using Q5 polymerase, using an annealing temperature of 59°C and an extension time of 45 seconds

with the remaining steps as per the manufacturer protocol. The amplified DNA was then purified using a Qiagen PCR purification kit and DNA concentration was measured using a Nanodrop spectrophotometer. The phosphorylation reaction was composed of 100 ng of purified linear DNA, adjusted to 10 µl using water, combined with 10 µl of 2X Quick Ligation buffer (NEB) and 1 µl of T4 polynucleotide kinase (NEB). This was allowed to incubate at 37°C for 30 mins and then cooled to room temperature for 5 mins. One µl of NEB Quick Ligase was then added and the reaction allowed to incubate at 25°C for another 15-30 min. Two µl of this reaction mixture was used to transform 50 µl of chemically competent *E. coli* DH5α or NEB Turbo cells and selected on plates with 100 µg/ml ampicillin.

Successful integration of the 20 bp protospacer sequence would replace one of three NcoI sites and thus can be first verified via restriction digestion and subsequently by sequencing using Primer set pSgRNA check.

4.2.6 Homologous Recombination

E. coli strain SW02 carrying plasmid pIRTCas9 was first transformed with plasmid pSgRNA-xxx by electroporation and selected for growth on chloramphenicol, ampicillin and IPTG containing plates. Large colonies were selected and grown at 37°C in LB medium supplemented with chloroamphenicol, IPTG and ampicillin to an optical density of 0.4 and λ*red* gene expression was induced by the addition of 0.4% arabinose for 30 minutes. Cells were harvested by centrifugation and then washed thrice with ice-cold 10% glycerol solutions. The cells were then resuspended in 1/100th of the original culture volume in 10% glycerol and aliquoted into 50 µl aliquots at -80ºC.

Donor oligonucleotides were added to 50 µl of cells at a final concentration of 10 µM and transferred to a 0.1 cm electroporation cuvette. Electroporation was performed on a BioRad Gene Pulser and cells were immediately recovered in 1 mL final volume of SOB with 1 mM IPTG for 45 min at 37ºC before plating on LB with 100 µg/ml anhydrotetracycline. Plates were incubated overnight at 37 °C or 30 °C for Ts strains. A small portion of the recovered cells were plated on LB as a control for transformation efficiency.

4.2.7 Plasmid Curing

The pSgRNA plasmid can be cured when manipulations are performed in a streptomycin resistant background. Strains were first grown in liquid culture and then subsequently plated on solid media, both supplemented with 300 µg/ml of streptomycin. Loss of the plasmid was verified by loss of ampicillin resistance. pITRCas9 is cured by growing in IPTG-free media for 10 or more doublings. Both plasmids can be cured simultaneously by growth in media containing streptomycin and lacking IPTG.

4.2.8 Efficiency of editing

The efficiency of the CRISPR/Cas9 system under different conditions is expressed as the ratio of the number of surviving cells in the absence and presence of anhydrotetracycline. All colony counts were calculated from serial dilutions and are the average of three independent trials.

4.3 EXPERIMENTAL RESULTS

4.3.1 Alignment of Compatible Origins of Replication

The pBR322, RSF1030 and p15a origins of replication are often used in cloning vectors. Though these plasmids all use the same unidirectional mechanism of replication dependent on an RNAII primer, they fall in different incompatibility groups and can hence be mainatained in the same cell. These origins have an additional benefit of having copy numbers ranging from 10 for p15a, 15-20 for pBR322 to 100 for RSF1030. Variants of pBR322 (namely pUC) or RSF1030 (namely RSF1031) can have very high copy numbers of 500-700 copies/cell. An alignment of the RNA pBR322, p15a and RSF1030 origins of replication allows us to map the RNAII priming region (Fig. 4.1). ColE1 is included in the alignment for comparison. The RNAI sequence, including its -10 and -35 promoter regions, are identical in all four replication origins. The RNAII sequences, differ between pBR322, p15a and RSF1030 which places them place them in different plasmid incompatibility groups. Note that ColE1 and pBR322 fall within the same group. The native -10 and -35 regions of the RNAII promoter were replaced with a Lac promoter under $LacI^{Q}$ control. Thus, RNAII production, and hence DNA replication, becomes dependent on induction by IPTG.

4.3.2 Relative Copy Numbers of Different Origins

pAM34 has a pBR322 origin of replication and is a 6 kb plasmid. Cronan[73] has shown that pAM34 has a slightly higher copy number (1.5 times) than a standard cloning vector with a pBR322 origin which has a literature value of 20. Plasmid pI1030 is observed to have a very similar copy number to pAM34 (Fig. 4.3) whereas the p15a origin of plasmid pI15 is only 60% of pAM34 giving it a copy number closer to that of a classical pBR322 plasmid like pBAD322S [73]. RSF1030 is a

medium copy origin (100 copies/cell). However, under *lacZp* control, significantly lower copies are seen. Moreover, the lacZ promoter has been previously shown to be fully induced at 1mM IPTG. Plasmid pI1031 has a single base change from pI1030 that is shown to increase copy numbers to pUC-like numbers (400-500). Though the copy number of pI1031 is roughly twice than pI1030 and pAM34, its copy number of 60 is lower than that of a classical RSF1030 origin.

Omitting ampicillin selection pressure, albeit during the first 6-7 generations, had very little influence on the copy number with a 5-10% depression in copy number observed. However, without IPTG, the copy numbers fall precipitously with pAM34 being the only plasmid still maintaining about 30% of its original copy number.

4.3.3 Dynamics of Curing of the pI series of Plasmids

pAM34 was shown to require IPTG concentrations of >500 nM for effective maintenance. By measuring plasmid yields from cultures grown in differing IPTG concentrations, all plasmids tested showed a 15-20% decrease in plasmid concentration when grown with only 0.1 mM IPTG (Fig 4.4). Since *E. coli* has an active lactose transporter, *lacY*, it is possible that higher intracellular concentrations are maintained. However, unlike plasmids pI15 and pI1030 which are efficiently lost when IPTG is withdrawn, pAM34 cultures retain 35-40% of their plasmid content.

When colonies grown with ampicillin and IPTG are streaked on an LB plate, almost all colonies were observed to have lost the plasmid. This was verified by screening for loss of ampicillin resistance and by attempting to purify plasmids from a small subset of colonies. Cultures carrying pI15, pI1030 and pI1031 completely lost their plasmids within 15 doublings, with pI15 and pI1031

losing more than 50% in less than 5 generations. pI1030, however, experiences complete plasmid loss in a similar time frame as the other two. Meanwhile, pAM34 requires more than 25 generations to get fully cured. The pAM34 curing rate agrees with similar observations made by Gill and Bouche [74], in the original report of pAM34.

4.3.4 Description and Validation of the Two-plasmid System in *E. coli*

To establish and improve the speed and utility of CRISPR/Cas based genome editing in *E. coli*, a two-plasmid system was constructed consisting of pITRCas9 and pSgRNA plasmids (Fig. 4.6). pITRCas9 encodes the S. pyogenes Cas9 endonuclease with an *ssrA* degradation under a tetracycline-inducible promoter, has the λ beta, gamma and exo genes under pBAD control and a p15a origin whose replication requires the presence of IPTG. The guide RNA plasmids encode a single guide RNA (pSgRNA) under a tetracycline promoter and contain an orthologous *rpsL* counterselection marker to allow for rapid curing of the plasmid in a streptomycin resistant strain. New spacers were introduced in pSgRNA using a PCR based cloning strategy. Alternatively, two Bbs1 sites were also used to introduce spacers using a golden gate approach (Fig. 4.6)

The effectiveness of the plasmids was tested by using guide RNA targeting the essential gene *dnaG* in *E. coli* BW25113 (Fig. 4.8). In the presence of anhydrotetracycline, a 10^4-fold reduction in colonies was observed. However, ~10^2 colonies escaped killing by Cas9.

4.3.5 Oligonucleotide-mediated Ts Mutations

CRISPR/Cas has been shown to be an effective scarless counterselection for generating point mutations using single stranded donor DNA. We demonstrate the particular utility of this tool in generating a point mutation in an essential gene cluster that results in a temperature sensitive yet plasmid free strain. Most genes in the fatty acid biosynthetic operon are essential and the most convenient way to study them has been using temperature sensitive variants that grow at lower temperatures but lose function at higher temperatures. Using pSgRNA-*fabG*, a single guide RNA that targets the 714th base of *fabG* (Table 4.3), and an oligonucleotide resembling the lagging strand, we successfully introduced a E233K mutation in *fabG* with an efficiency of 99% (Table 4.4).

4.3.6 Short Deletions

Expression of the λ*red* genes has been successfully used in the past for ssDNA mediated deletion up to 45 kb of the genome. We used a sgRNA and an oligonucleotide to target and delete a short 32 bp sequence from the *fabG* gene that corresponds to an 8 residue in-frame deletion in the gene. Of the 72 colonies tested, 94% had the deletion (Table 4.4). This was first tested by temperature sensitive growth and then by sequencing.

4.3.7 Short Sequence Insertions

We tested the ability of this system to insert short sequences into the *E. coli* genome. λRed-mediated recombination has been used to make both large and small insertions using just 50 bp regions of homology using donor dsDNA. However, to introduce short tags like the hexahistadine tag (6X-His), ssDNA has been used to insert up to 30 bp. We use a 95 bp ssDNA to insert a 24 bp

hexahistidine tag with a Gly-Gly linker at the C-terminus of *fabG*. We obtained mutants at an efficiency of 90% (Table 4.4). Colonies were first tested for the insertion via PCR and subsequently confirmed by sequencing. Nine colonies were tested.

4.3.8 Speed and Efficiency of Plasmid Curing

Due to the limited availability of compatible origins of replication and commonly used genetic markers for strain manipulations, recombineering methods resulting in strains free of plasmids are generally preferred. The pITRCas9 construct is based on the pI15 vector backbone. Vector pI15 was shown to be cured within 12-15 generations. Given the large payload and hence the size of pITRCas9, the copy number is expected to be much lower. pITRCas9 can hence be completely cured in less than 6-7 generations which is easily achievable in a single culture grown to saturation. Meanwhile, pSgRNA can be cured by growing the strain in liquid culture in the presence of streptomycin (300 μg/ml) and then plated on LB agar plates with the same streptomycin concentration. The two independent curing mechanisms can be utilized to facilitate either individual or simultaneous loss of both plasmids.

4.4 DISCUSSION

CRISPR/Cas systems were originally identified for their role in bacterial adaptive immunity but have recently taken the scientific world by storm due to their remarkable ability to target complex genomes with just a small guide RNA. In prokaryotes, recombineering using phage derived recombination machinery remains one of the most effective methods of genetic engineering. However, the major limiting aspect of this remains the process of selecting mutants over wild-type

cells. The combination of recombineering with CRISPR/Cas9 based counter-selection systems has greatly enhanced the effectiveness of this selection process and have since been incorporated into a number of tools for microorganisms into which plasmids can be introduced like *E. coli*, *S. pneumoniae*, *L. reuteri* and *T. citrea* amongst others. Though single-plasmid systems containing the guide RNA, the endonuclease and/or the donor DNA have been devised for other bacteria and eukaryotes, this is not possible in the specific case of *E. coli* if it also has to be used as a cloning host. Jiang et al first demonstrate its utility in *E. coli* with two plasmids, one encoding the spCas9 and the other with a CRISPR array targeting a gene of interest. This arrangement requires expression of the λ*red* genes, either from the genome or on a third helper plasmid in order to use ssDNA or short dsDNA as recombination templates. Further iterations have combined different combinations of Cas9, λ*red* and guide RNA (sgRNA or crRNA) in a two-plasmid format under either constitutive or inducible promoters. Recently reported tools also utilize temperature sensitive origins of replication to facilitate guide RNA plasmid curing at higher temperatures but still require specific efforts to cure the Cas9 encoding plasmid, costing time in iterative editing efforts.

In this Chapter, I present a convenient and facile two-plasmid system. The first plasmid (pITRCas9) encodes Sp. Cas9 under a tetracycline inducible promoter, the λ*red* genes under an arabinose inducible expression and an IPTG-dependent origin of replication. The second plasmid contains a single guide RNA (pSgRNA) also under tetracycline control with an *rpsL* counterselection system. This combination provides some unique advantages and facilitates ease of use. The pSgRNA plasmid is small (~4 kb) and new protospacers can be introduced via a single PCR reaction followed by circularization and transformation into a cloning host like DH5α. The

empty vector also has BbsI sites that maybe used with short annealed oligos for Golden Gate cloning. Multiple spacers may also be cloned into either vector using synthesized DNA blocks and/or Gibson assembly to simultaneously target multiple genes. Since the plasmids can be independently cured, loss of either one or both plasmids can be achieved in a single step depending on whether pITRCas9 is required for subsequent edits or not. Also, since this may be performed independent of any strain specific temperature requirement, cloning, growth, recombination and curing may be performed at the any temperature optimal for the experimental design.

A typical workflow for introducing a single change takes about 4 days from start to finish which is almost twice as fast than any currently available system to obtain plasmid free strains. Further, changes may be achieved in half the time. The *rpsL* counterselection system uses a wild-type *rpsL* gene from *Photorhabdus luminescens* that confers a dominant sensitivity to streptomycin in a streptomycin resistant background. The sequence of the *P. luminescens rpsL* gene has been optimized for use in *E. coli* and has been verified to have no significant stretch of homology to the endogenous *rpsL* gene. The only requirement for *rpsL* based counterselection is a streptomycin resistant background that is easily achieved by either using some popular strains such as Top10, MC4100, MC1061, JM103, JM105, JM83, EPI300, DH10B, AB1157 and their derivatives or by simply transforming a short oligo to introduce a point mutation in the wild type *rpsL* gene and selecting for streptomycin resistance, as shown in BW25113 to generate SW02.

Most escapers observed in *E. coli* maybe attributed to a mutation in either the Cas9 nuclease or the spacer sequence. This is an exploding field of study and more efficient Cas9 nucleases are being discovered/developed. Given its size, the *cas9* gene is a large target for mutations. In a scenario

where iterative changes are required, this can prove to be cumbersome since curing the Cas9 encoding plasmid and swapping it out for a new one might be rate-limiting. This plasmid can be potentially shortened by including the repressive elements such as $LacI^{Q}$, TetR and AraC on the host genome. However, the biggest improvement might come from discovery of minimal or smaller Cas9 nucleases.

Plasmid pITRCas9 is itself based on pI15 which has a IPTG-dependent p15a origin of replication. p15a has a low copy number of 10-15 and thus might be useful to reduce any potential toxicity that Cas9 might have prior to induction. The differing copy numbers provided by pI1030, pI1031 or pAM34 might be useful to regulate the expression of other Cas9 nucleases

4.5 TABLES AND FIGURES

Strains	Relevant Genotype	Source
MC1061	Wild type strain	Casadaban, et al., 1980
DH5α	*endA1 recA1 0 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17*	Hanahan, et al., 1985
BW25113	*lacI+rrnBT14 ΔlacZWJ16 hsdR514 ΔaraBADAH33 ΔrhaBADLD78 rph-1*	CGSC #7636
SW02	*BW25113 rpsL*135 (strR)	This work

Table 4.1 Bacterial Strains used in this study

Plasmids	Relevant Characteristics	Source
pAM34	IPTG-dependent plasmid with pBR322 *ori*	Gill et. al, 1991
pACYC184	Cloning vector with p15a *ori*	Chang et. al, 1978
pCY2016	pBAD plasmid with RSF1030 *ori*; medium copy no	Chakravartty et al, 2015
pCY2017	pBAD plasmid with RSF1031 *ori* i; high copy no	Chakravartty et al, 2015
pI15	IPTG-dependent plasmid with p15a *ori*	This work
pI1030	IPTG-dependent plasmid with RSF1030 *ori*	This work
pI1031	IPTG-dependent plasmid with RSF1031 *ori*	This work
pKD46	Amp^R, Ts ori, Arabinose-inducible expression of λ*red* genes	Datsenko and Wanner, 2000
pCas9CR4	Cas9 nuclease under pTet promoter and constitutive TetR	Reisch et al, 2015
pKDsgRNA-p15	pSC101 (Ts) origin; Tet-inducible protospacer targeting the p15a *ori*, Ara-inducible λ*red*, $Spec^R$	Reisch et. al, 2015
pSW09	pCas9CR4 with IPTG-dependent replication	This work
pSW10	pKDsgRNA-p15a w/o λ*red*, pSC101 Ts, Amp^R	This work
pSW11	pSW10 non-Ts, Amp^R	This work
pITRCas9	IPTG-dependent replication, Ara-inducible λ*red*, Cas9 under pTet promoter; Cm^R	This work
pSgRNA	pSC101 (non-Ts); Amp^R; protospacer sequence with Bbs1 sites, Tet-inducible guide RNA production	This work
pSgRNA-p15a	Protospacer targeting p15a origin; ctatcgtcttgagtccaacc	This work
pSgRNA-*dnaG*	Protospacer targeting *dnaG*; ggagctctggacattaaacc	This work
pSgRNA-*fabG*	Protospacer targeting *fabG*; ggtgaaactttgcatgtgaa	This work
pSgRNA-*fabG2*	Protospacer targeting *fabG*; tgtgaacggcgggatgtaca	This work

Table 4.2 Plasmids used in this study

Oligonucleotides	Sequence
p15a Kpn1	ATGCCAGGTACCTTGCTCTGAAAACGAAAAAACCGC
p15a Xba1	GATCCATCTAGAACTAGAGTCACACTGGCTCAC
RSF Kpn1	ATGCCAGGTACCTCATAAATAAAGAAAAACCACCGC
RSF Xba1	GATCCATCTAGACTCTTTTGTTTATTTTTCTAAATACATTC
p34C AatII F	ATCGAGACGTCTTGAAATAAGATCACTACCGGG
p34C Xba1 R	ATGCCATCTAGAGGATCCGCATGCCTGCAGGTCGACAAGCTTC CCGGGTACCGAGCTCGCGAATTTCTGCCATTCATC
p34K AatII F	ATCGAGACGTCCCACGTTGTGTCTCAAAATCTCTG
p34K Xba1 R	ATGCCATCTAGAGGATCCGCATGCCTGCAGGTCGACAAGCTTC CCGGGTACCGAGCTCCGCTGAGGTCTGCCTCG
p34G AatII F	ATCGAGACGTCGAATTGACATAAGCCTGTTCGGTTCG
p34G Xba1 R	ATGCCATCTAGAGGATCCGCATGCCTGCAGGTCGACAAGCTTC CCGGGTACCGAGCTCGAATTGGCCGCGGCGT
p15a-lacI Spe1 F	TCTAGAACTAGTGAAACCATTATTATCATGACATTAACC
p15a-lacI Age1 R	CGGTTTACCGGTGTCATTC
LRed AatII F	ATGCCAGACGTCCATCGATTTATTATGACAACTTGACGG
LRed term AatII R	GATCAGGACGTCAAGCTTAAGTCAAAAGCCTCCGACCGGAGGC TTTTGACTATTACCATGGTCATCGCCATTGCTCCC
Amp Nco1 F	ATGACACCATGGTAAATACATTCAAATATGTATCCGCTC
Amp Xho1 R	GCATGACTCGAGGTAAACTTGGTCTGACAGTTAC
Non Ts F	GCTTACTTTGCATGTCACTC
Non Ts R	ATGATCTCAATGGTTCGTTC
pSgRNA-*dnaG* F	GGAGCTCTGGACATTAAACCGTTTTAGAGCTAGAAATAGCAAG
pSgRNA-*fabG* F	GGTGAAACTTTGCATGTGAAGTTTTAGAGCTAGAAATAGCAAG
pSgRNA-*fabG2* F	TGTGAACGGCGGGATGTACAGTTTTAGAGCTAGAAATAGCAAG
pSgRNA R	GTGCTCAGTATCTCTATCACTGA
pSgRNA check	GCTACGCCTGAATAAGTG

Table 4.3 Oligonucleotides used in this study

Target	**Mutation**	**CFU**	**Positives**
Control	None	5.2×10^7	N.a.
dnaG	None	2.7×10^3	N.a.
fabG	Ts (point)	4.5×10^5	99 ± 1%
fabG	24bp deletion	1.7×10^4	94 ± 3%

Table 4.4 Efficiency of CRISPR/Cas9 system

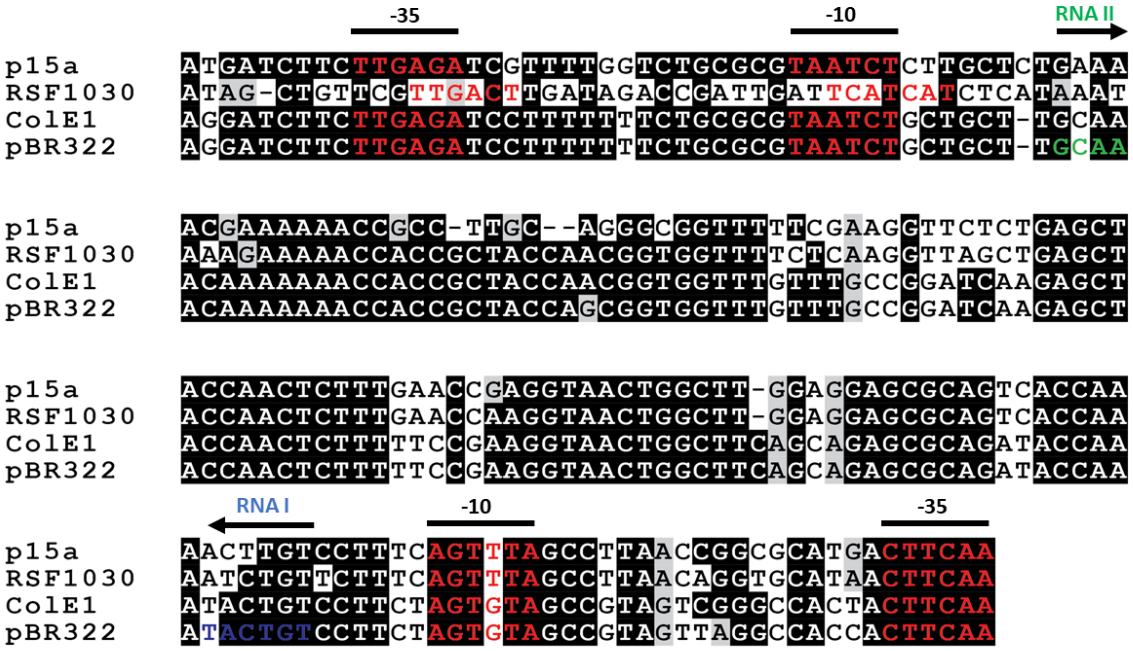

Figure 4.1 Comparison of RNA I and RNA II binding sequences of different replication origins. Alignment was performed with ClustalW. Conserved sequences are highlighted in black. The -10 and -35 boxes of the RNA II and RNA I are indicated in red. Known RNA II primer sequence of pBR322 is indicated in green while the RNA I sequence is in blue. Strong conservancy is observed between the -10 and -35 boxes of p15a, pBR322 and ColE1. However, the RNA II primer sequences differ enough to place them in different incompatibility groups.

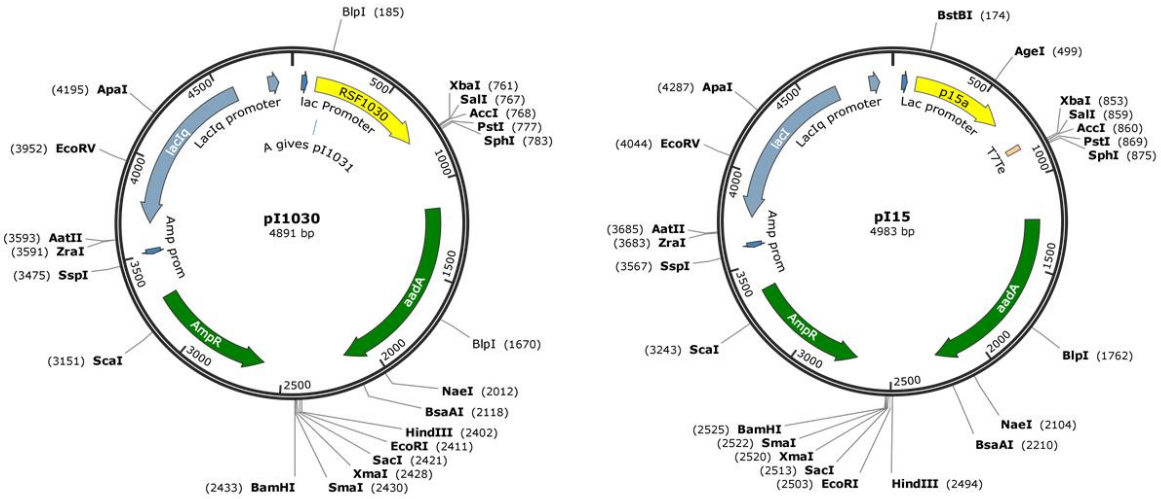

Figure 4.2A Plasmid maps of pI15 and pI1030. pI15 and pI1030 are smaller than pAm34 due to deletion of *rop* along with the pBR322 origin. *rop* is shown to not function with the p15a origin of replication. Other elements including the multiple cloning site (MCS), *aadA* gene and the direct repeat regions found in pAM34 are intact. Antibiotic resistance markers are in green and replication origins are in yellow.

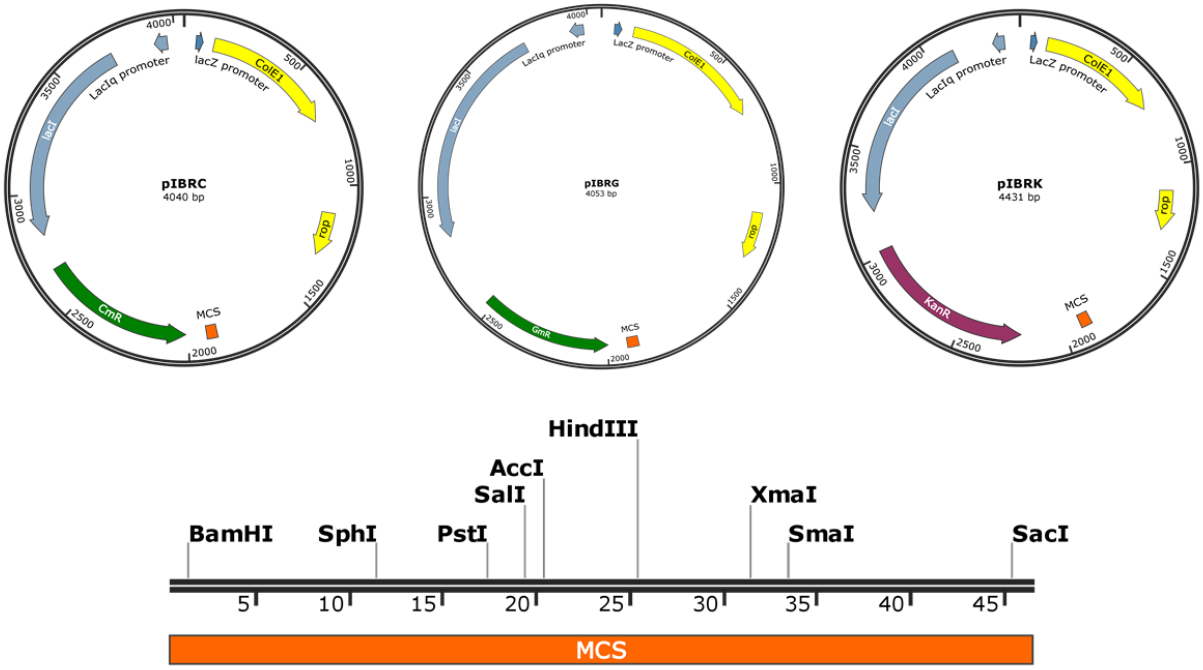

Figure 4.2B Plasmid maps of pIBRC, pIBRK and pIBRG are shown. These plasmids are much smaller than pAM34 and are resistant to chloramphenicol, kanamycin and gentamycin respectively. They are missing the *aadA* gene and the old MCS. The restriction sites present in the new MCS are shown. Note: BamH1 is duplicated and the two sites flank Xba1. AccI is not unique as well.

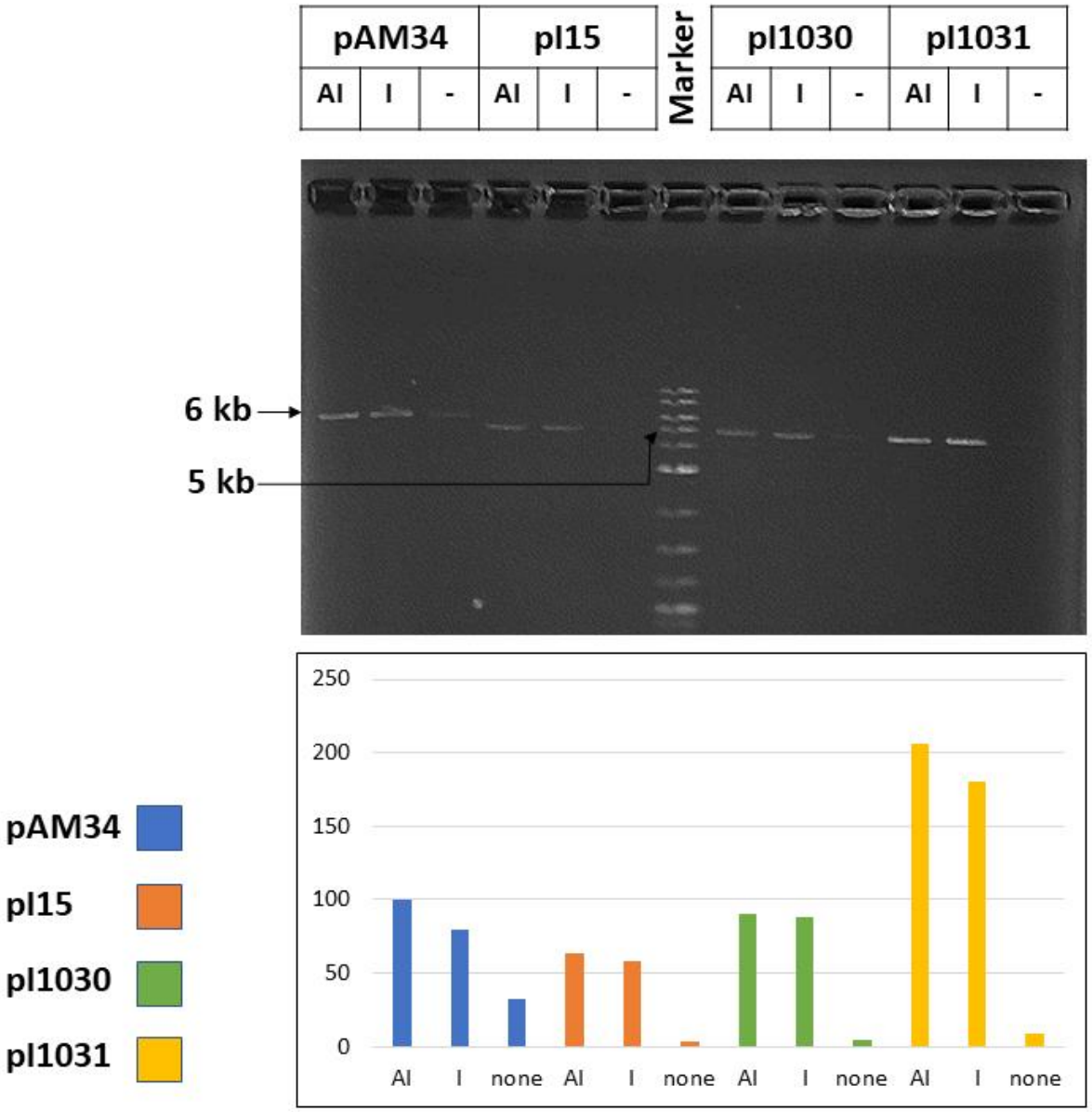

Figure 4.3 IPTG-dependent replication of pAM34, pI15, pI1030 and pI1031. Plasmids were extracted from 1ml of an LB culture grown to OD 1.0 from OD 0.01 with the following supplements. AI represents addition of ampicillin (100ug/ml) and IPTG (1mM), I represents addition of IPTG (1 mM) only while – represents no supplementation. Relative plasmid amounts are indicated in the bar graph below with pAM34 set as 100. All four strains show plasmid loss in the absence of IPTG while the lack of ampicillin selection has minimal to no effect in 6-7 generations. Marker: NEB 2-log ladder.

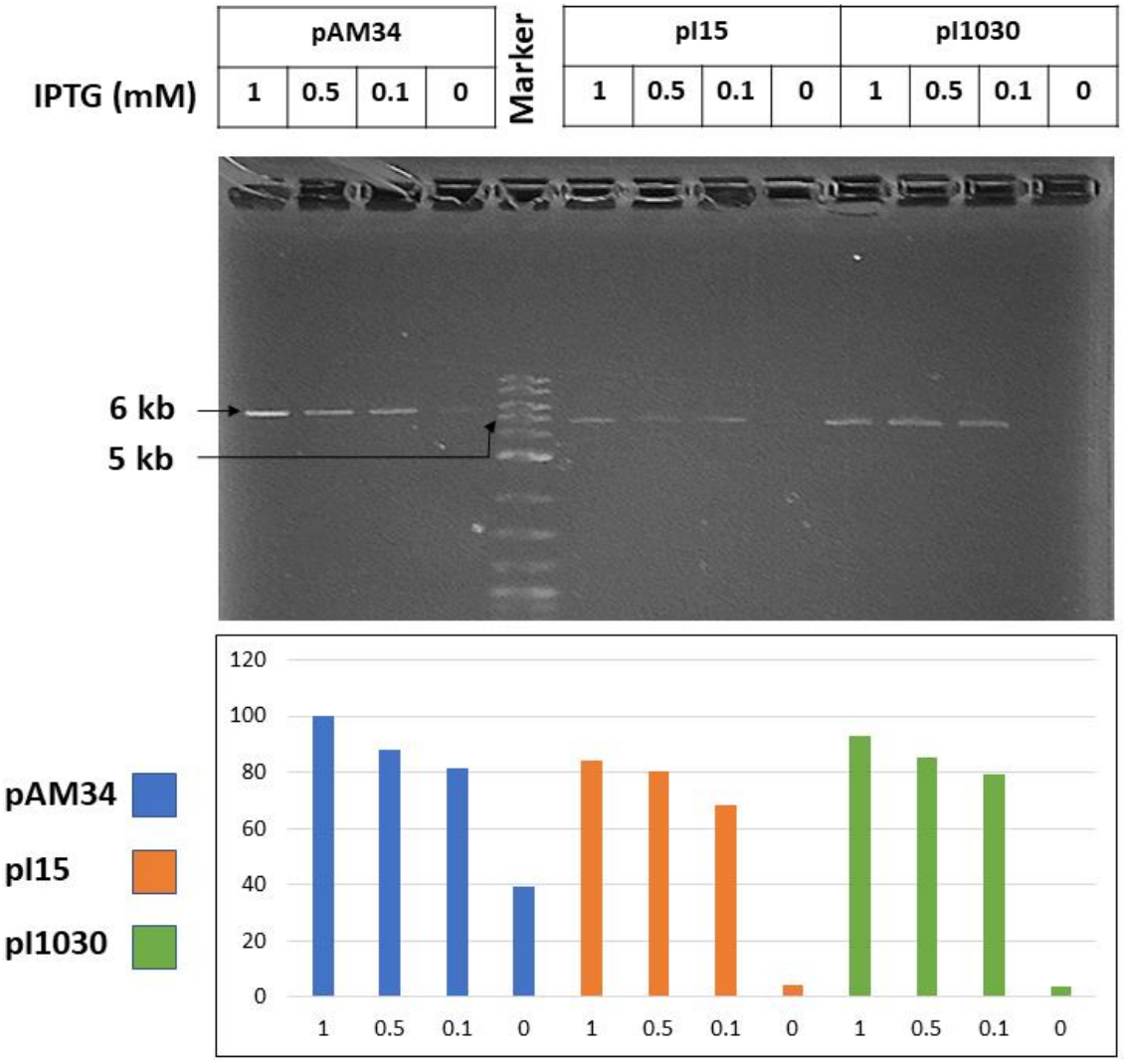

Figure 4.4 Changes in plasmid levels with varying IPTG concentrations. Plasmids were extracted from 1 ml of an LB culture grown to OD 1.0 from OD 0.01. Cultures were supplemented with varying IPTG amounts as indicated above each lane. Relative plasmid amounts are indicated in the bar graph below with pAM34 set as 100. All four strains show plasmid loss in the absence of IPTG while the lack of ampicillin selection has minimal to no effect in 6-7 generations. Marker: NEB 2-log ladder.

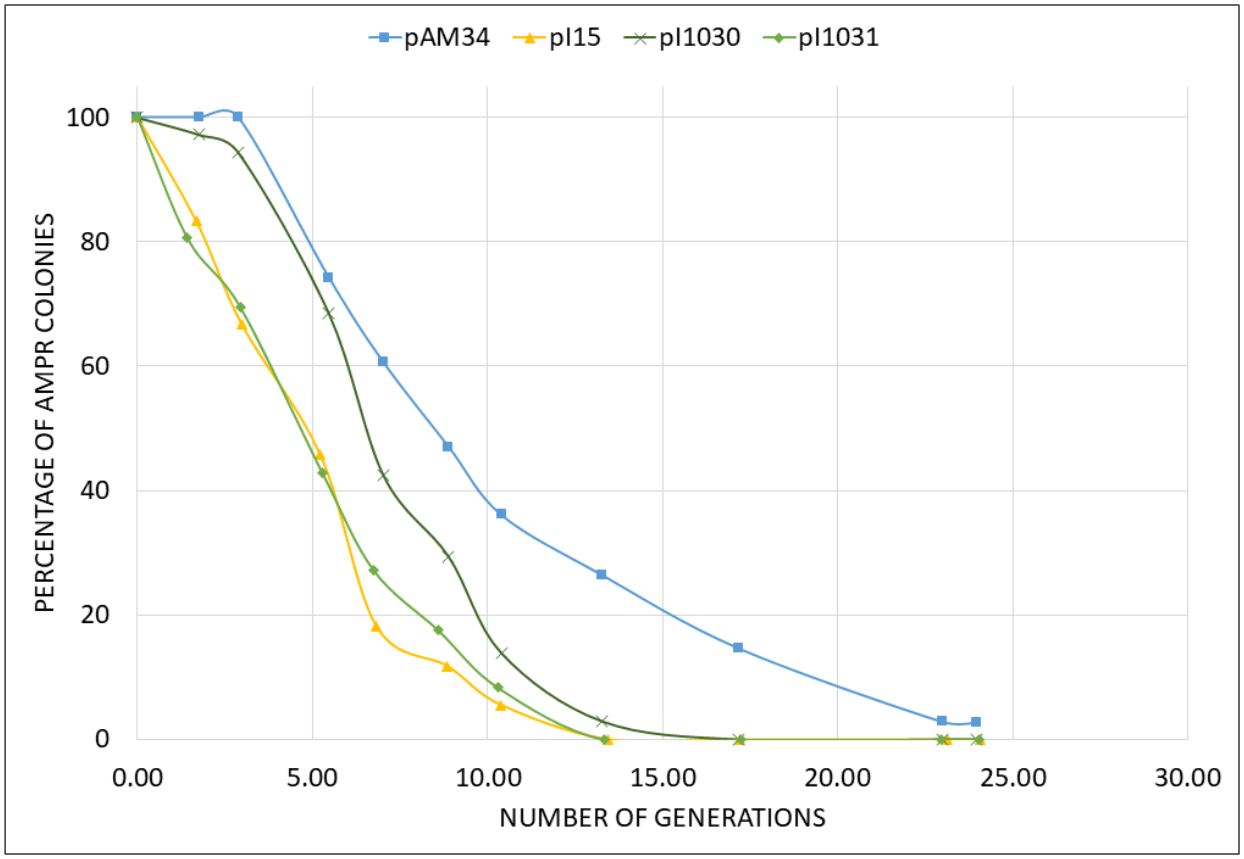

Figure 4.5 Rate of curing of pAM34, pI15, pI1030 and pI1031. Plasmid curing was measured over 25 generations in LB with no IPTG supplementation. Cultures were started at OD 0.01 and samples were withdrawn periodically and plated on LB agar with 1 mM IPTG. OD measurements were used to calculate number of generations grown. Samples were re-diluted to OD 0.01 when an OD of 1.0 was achieved. Ratio of ampicillin resistant colonies to total number of colonies is represented against number of generation. Thirty six colonies were analyzed for each time-point. Strains are cured of pI15, pI1030 and pI1031 within 15 generations and at a significantly faster rate than pAM34 which was not completely cured even after 25 generations. Hence, pI1030 and pI1031 are more unstable than pI15.

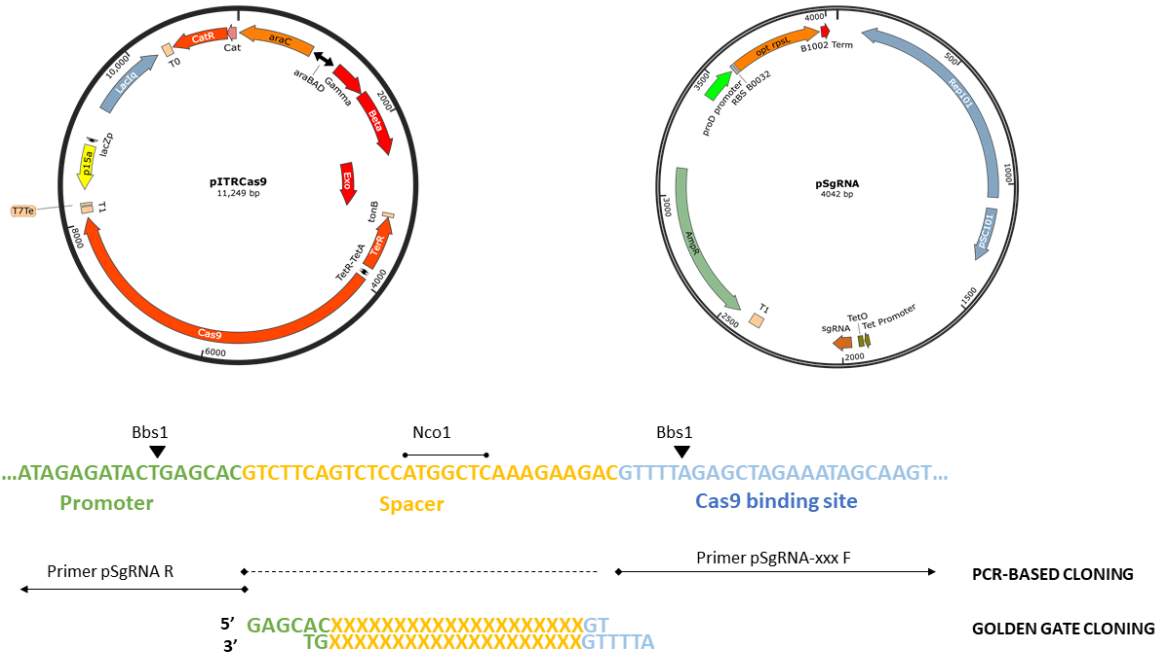

Figure 4.6 Plasmid maps of the CRISPR/Cas9 vectors. Plasmid maps show key features of both plasmids. pITRCas9 contains *cas9* under a tetracycline-inducible promoter. The tetracycline repressor is expressed from its own promoter. The lambda RED genes are under arabinose control with constitutive expression of AraC from its own promoter. Replication is from a p15a origin. The RNA II primer is under a *lacZ* promoter and hence replication is IPTG-dependent. pITRCas9 confers resistance to chloramphenicol. pSgRNA as a non-temperature sensitive pSC101 origin and is confers ampicillin resistance. It contains a tetracycline inducible single-guide RNA. The empty vector has a protospacer (depicted below the map) that has Bbs1 sites for golden gate cloning and a Nco1 site (1 of 3) for quickly verifying protospacer insertion. Two strategies for cloning are depicted below: a PCR based approach that includes the protospacer sequence on one primer and a golden gate approach that requires specific overhangs included with the protospacer sequence (green and blue) during oligo synthesis.

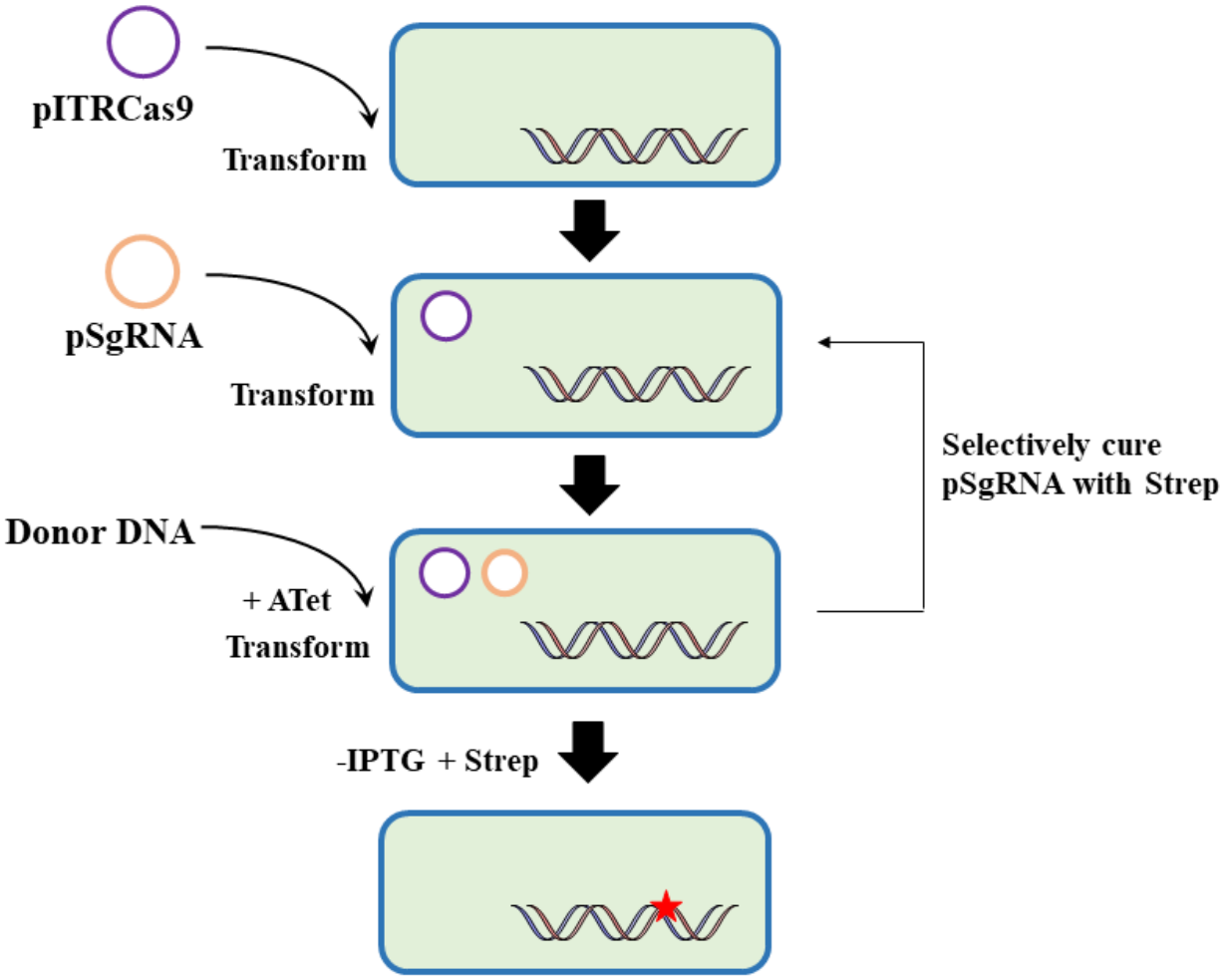

Figure 4.7 Schematic for single/iterative knockouts. pITRCas9 and pSgRNA are sequentially transformed into the host strain. Lambda red is induced during competent cell preparation and 10 mM of donor DNA is transformed. IPTG is added during recovery but can be omitted while plating to induce curing of pITRCas9. For iterative changes, pSgRNA can be selectively cured by growing on streptomycin containing media (provided the host is streptomycin resistant) and transforming a new pSgRNA.

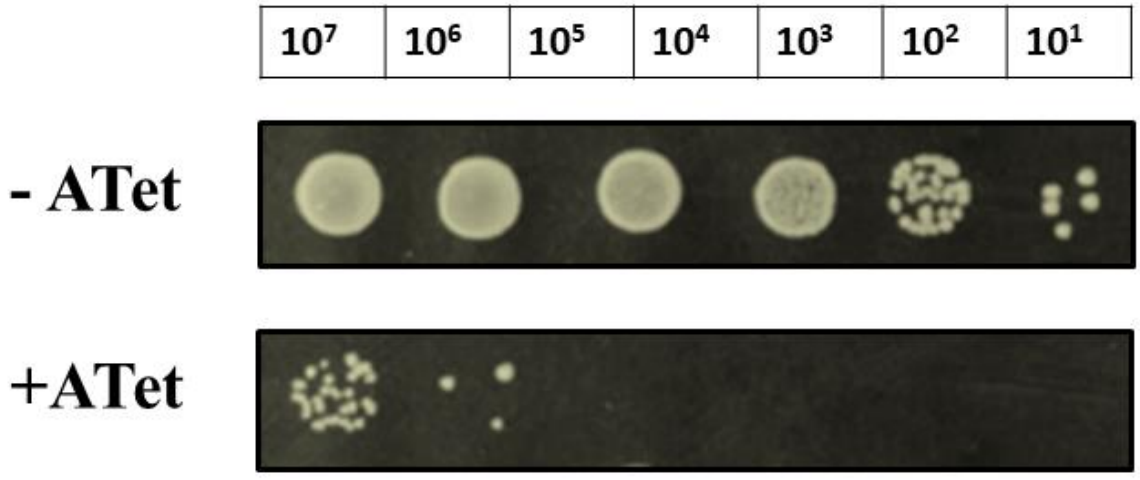

Figure 4.8 Efficiency of killing. Ten ul of a ten-fold serial dilution of strain BW25113 containing pITRCas9 and pSgRNA-*dnaG* was plated on LB agar plates with or without 100 ug/ml anhydrotetracycline. Greater than 10^4 killing is observed upon induction of the CRISPR-Cas9 system

Chapter Five

CONCLUSIONS

Fatty acids are an essential component of living organisms. Though they serve multiple roles both inside and outside the cell, in bacteria their primary purpose is to build the cellular membranes. The composition of these membranes is carefully regulated to maintain its functionality under various physiological and stress conditions. Since fatty acid synthesis is energetically expensive, these processes are regulated at multiple levels. Individual examples of this in *E. coli* include the regulation of acetyl-CoA carboxylase by phosphorylation and allostery and transcriptional regulation of unsaturated fatty acid synthesis by the repressor FabR. Other organisms which can shuttle intermediates between their fatty acid biosynthetic and degradation pathways show even more complex regulation. A holistic picture of this regulation is unavailable because of a lack of clarity regarding the composition of intermediates at any given condition.

In Chapter 1, I present a rapid and facile method to isolate and characterize the ACP bound intermediates from *E. coli* and *B. subtilis*. This method requires a single genetic change to insert an inert Strep-tag at the carboxy-terminus of ACP on the chromosome. In *E. coli* and *B. subtilis*, this change has no effect on the growth of the organism or the functionality of the ACP protein. The ACP isolation steps are otherwise unaffected by any other genetic change or treatment being tested. A significant amount of acetyl-ACP was discovered in both *E. coli* and *B. subtilis.* This raises certain pressing questions such as whether acetyl-ACP is metabolically active and, if so, why such a significant amount of carbon is dedicated to this intermediate. Given the non-essentiality of *fabH* and the detection of acetyl-ACP, a robust bypass pathway probably exists in

E. coli that can condense acetyl-ACP with malonyl-ACP. Discovering this enzyme might not be trivial because no other gene is found to be overexpressed in *fabH* mutants in comparison to the wildtype. Observing changes in the acetyl-ACP pool in *fabH*, *fabD*, *fabB* and *fabF* mutants might provide a clue. Further, the use of fatty acid initiation inhibitors like cerulenin and thiolactomycin might also demonstrate how this metabolic pool varies.

We observe no other dicarboxylic acyl-ACPs other than malonyl-ACP in both *E. coli* and *B. subtilis*. The biotin biosynthetic pathway in *B. subtilis* is complicated and not completely elucidated. Though we know that *B. subtilis* uses a free pimelic acid precursor that is directly or indirectly synthesized through the fatty acid pathway, the source of this remains a mystery. Biotin auxotrophic mutants in *B. subtilis* might accumulate certain early precursors providing a clue to drive further research in this area.

The current view of bacterial fatty acid synthesis is still quite *E. coli* centric, and numerous examples exist of other bacteria following 'new rules'. The abundance of ACP in *E. coli* has given us the false perception that ACP itself cannot be a limiting factor in fatty acid related processes. I show through multiple independent approaches that *B. subtilis* also has only a tenth of the ACP present in *E. coli*. Preliminary evidence from other Gram-positive and Gram-negative bacteria indicates that this paucity might be more prevalent. In *B. subtilis,* ACP associates very tightly with FabF, with this interaction persisting through multiple purification columns. Perhaps this tight interaction is part of a larger loose complex of fatty acid enzymes that effectively increases the local availability of ACP to its cognate enzymes. Current efforts are focusing on seeing if the

reciprocal interaction can be observed wherein ACP (and perhaps the other fatty acid synthesis enzymes) co-elute(s) during FabF purification.

Meanwhile, one of the most direct utilities for the method remains for system biologists and synthetic biologists. An acyl-ACP snapshot would allow us to trace metabolic fluxes through the fatty acid pathway in an unprecedented manner. This would allow them to fine tune metabolic models to accommodate their observations. This is especially useful when used in an industrial organism wherein productivity can be rationally improved.

In Chapter 3, I characterized a novel temperature-sensitive mutant of *E. coli* FabG and demonstrated this temperature-sensitivity to be due to the loss of an eight-amino acid section that consequently reduced dimerization. By mutating the wildtype copy of this gene on the *E. coli* genome, its effects on phospholipid synthesis and lipid A synthesis could be investigated. Interestingly, the defects in this mutant could be compensated by simultaneously mutating a gene in lipid A synthesis, *lpxC*, suggesting that this is a point of regulation between these two pathways towards balanced growth. The mutant also, surprisingly, reverted to temperature-resistance with the frequency of a point mutant. Sequencing these revertants showed that this was caused by other compensatory mutations in the dimer interface. An exhaustive search for all compensating mutations was not performed. Almost no drugs that target FabG exist, even though FabG homologs are found in almost all bacteria and they all function as multimers. Identifying and characterizing the suppressive mutants might contribute to a better understanding of the specific interactions required to stabilize the FabG dimer/tetramer and help guide rational drug design.

In Chapter 4, I describe the development of a CRISPR/Cas9 toolkit along with a new set of vectors that have facilitated studies in the previous chapters. These are conditional vectors that can replicate only in the presence of IPTG. These plasmids utilize origins of replication that all require the production of a RNA II primer that binds to the origin and initiates DNA replication. Transcription of this RNA primer has been placed under a *lacZ* promoter which is repressed by $LacI^{Q}$ in the absence of IPTG. This has been extended to other compatible plasmids that have a RNA II-based mechanism of initiation but with different primer sequences. These plasmids can be reliably cured within 15 generations with more than 50% of plasmid loss occurring within the first five generations.

A CRISPR/Cas9 vector toolkit has also been developed wherein the expression of the Cas9 nuclease is under a tightly regulated tetracycline promoter while the replication of the plasmid itself is IPTG-dependent and is built off the previously described plasmids. This allows this plasmid to be rapidly cured from strains where the genetic manipulations are complete. The system is effective and can provide a 10^4-fold selection for recombination. This is comparable with other CRISPR/Cas9 systems. The second plasmid contains a single guide RNA also under tetracycline control and a *rpsL* counter-selectable marker. Wildtype *rpsL* confers dominant streptomycin sensitivity. In a streptomycin resistant background strain, plasmid loss can be selected for by growing the plasmid carrying strain in the presence of streptomycin. Another approach being explored is expressing a chromophore, wherein strains which have lost a plasmid can be immediately identified by loss of color. Interestingly, both plasmids can co-exist in the cell without inducing a double strand break if they are transformed sequentially and in the absence of inducer. This suggests that it might be theoretically possible to incorporate both the guide RNA and Cas9

on the same plasmid under the same tetracycline promoter, provided the tetracycline repressor is provided in *trans*. This possibility remains to be investigated.

REFERENCES

1 Maier, T. Architecture of Mammalian Fatty Acid Synthase at 4.5 A Resolution. *Science* **311**, 1258-1262, doi:10.1126/science.1123248 (2006).

2 Maier, T., Leibundgut, M. & Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **321**, 1315-1322, doi:10.1126/science.1161269 (2008).

3 Jenni, S., Leibundgut, M., Maier, T. & Ban, N. (Protein Data Bank, Rutgers University, 2006).

4 Cronan, J. E. & Waldrop, G. L. Multi-subunit acetyl-CoA carboxylases. *Progress in Lipid Research* **41**, 407-435, doi:10.1016/s0163-7827(02)00007-3 (2002).

5 Joshi, V. C. Mechanism of malonyl-coenzyme A-acyl-carrier protein transacylase. *Biochemical Journal* **128**, 43P.42-44P, doi:10.1042/bj1280043pb (1972).

6 Lowe, P. N. & Rhodes, S. Purification and characterization of [acyl-carrier-protein] acetyltransferase fromEscherichia coli. *Biochemical Journal* **250**, 789-796, doi:10.1042/bj2500789 (1988).

7 Weeks, G. & Wakil, S. J. in *Methods in Enzymology* 66-73 (Elsevier, 1969).

8 Rawlings, M. & Cronan, J. E. The Gene Encoding Escherichia-Coli Acyl Carrier Protein Lies within a Cluster of Fatty-Acid Biosynthetic Genes. *J Biol Chem* **267**, 5751-5754 (1992).

9 Heath, R. J. & Rock, C. O. Roles of the FabA and FabZ beta-hydroxyacyl-acyl carrier protein dehydratases in Escherichia coli fatty acid biosynthesis. *J Biol Chem* **271**, 27795-27801 (1996).

10 Garwin, J. L., Klages, A. L. & Cronan, J. E., Jr. Structural, enzymatic, and genetic studies of beta-ketoacyl-acyl carrier protein synthases I and II of Escherichia coli. *J Biol Chem* **255**, 11949-11956 (1980).

11 Bi, H., Zhu, L., Jia, J., Zeng, L. & Cronan, J. E. Unsaturated Fatty Acid Synthesis in the Gastric Pathogen Helicobacter pylori Proceeds via a Backtracking Mechanism. *Cell Chem Biol* **23**, 1480-1489, doi:10.1016/j.chembiol.2016.10.007 (2016).

12 Pramanik, J. & Keasling, J. D. Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol Bioeng* **56**, 398-421, doi:10.1002/(SICI)1097-0290(19971120)56:4<398::AID-BIT6>3.0.CO;2-J (1997).

13 Christensen, Q. H. & Cronan, J. E. Lipoic acid synthesis: a new family of octanoyltransferases generally annotated as lipoate protein ligases. *Biochemistry* **49**, 10024-10036, doi:10.1021/bi101215f (2010).

14 Walsh, C. T., Gehring, A. M., Weinreb, P. H., Quadri, L. E. N. & Flugel, R. S. Post-translational modification of polyketide and nonribosomal peptide synthases. *Current Opinion in Chemical Biology* **1**, 309-315, doi:10.1016/s1367-5931(97)80067-1 (1997).

15 Mofid, M. R., Finking, R. & Marahiel, M. A. Recognition of Hybrid Peptidyl Carrier Proteins/Acyl Carrier Proteins in Nonribosomal Peptide Synthetase Modules by the 4'-Phophopantetheinyl Transferases AcpS and Sfp. *J Biol Chem* **277**, 17023-17031, doi:10.1074/jbc.m200120200 (2002).

16 Roujeinikova, A. *et al.* X-ray crystallographic studies on butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site. *Structure* **10**, 825-835 (2002).

17 Gully, D., Moinier, D., Loiseau, L. & Bouveret, E. New partners of acyl carrier protein detected in Escherichia coli by tandem affinity purification. *FEBS Lett* **548**, 90-96 (2003).

18 Babu, M. *et al.* Structure of a SLC26 anion transporter STAS domain in complex with acyl carrier protein: implications for E. coli YchM in fatty acid metabolism. *Structure* **18**, 1450-1462, doi:10.1016/j.str.2010.08.015 (2010).

19 Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature* **433**, 531-537, doi:10.1038/nature03239 (2005).

20 Thomas, J. & Cronan, J. E. The enigmatic acyl carrier protein phosphodiesterase of Escherichia coli: genetic and enzymological characterization. *J Biol Chem* **280**, 34675-34683, doi:10.1074/jbc.M50573620010.1074/jbc.M505736200. Epub 2005 Aug 17. (2005).

21 Bligh, E. G. & Dyer, W. J. A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology* **37**, 911-917 (1959).

22 Agarwal, V., Lin, S., Lukk, T., Nair, S. K. & Cronan, J. E. Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis. *Proc Natl Acad Sci U S A* **109**, 17406-17411, doi:10.1073/pnas.1207028109 (2012).

23 Manandhar, M. & Cronan, J. E. A Canonical Biotin Synthesis Enzyme, 8-Amino-7-Oxononanoate Synthase (BioF), Utilizes Different Acyl Chain Donors in Bacillus subtilis and Escherichia coli. *Appl Environ Microbiol* **84**, doi:10.1128/AEM.02084-17 (2018).

24 Meier, J. L., Haushalter, R. W. & Burkart, M. D. A mechanism based protein crosslinker for acyl carrier protein dehydratases. *Bioorg Med Chem Lett* **20**, 4936-4939, doi:10.1016/j.bmcl.2010.06.028 (2010).

25 Bruegger, J. *et al.* Probing the selectivity and protein.protein interactions of a nonreducing fungal polyketide synthase using mechanism-based crosslinkers. *Chem Biol* **20**, 1135-1146, doi:10.1016/j.chembiol.2013.07.012 (2013).

26 Manandhar, M. & Cronan, J. E. Pimelic acid, the first precursor of the Bacillus subtilis biotin synthesis pathway, exists as the free acid and is assembled by fatty acid synthesis. *Mol Microbiol* **104**, 595-607, doi:10.1111/mmi.13648 (2017).

27 Rock, C. O., Garwin, J. L. & Cronan, J. E., Jr. Preparative enzymatic synthesis of acyl-acyl carrier protein. *Methods Enzymol* **72**, 397-403 (1981).

28 Post-Beittenmiller, D., Jaworski, J. & Ohlrogge, J. In vivo pools of free and acylated acyl carrier proteins in spinach. Evidence for sites of regulation of fatty acid biosynthesis. *J Biol Chem* **266**, 1858-1865 (1991).

29 Hicks, L. M., O'Connor, S. E., Mazur, M. T., Walsh, C. T. & Kelleher, N. L. Mass spectrometric interrogation of thioester-bound intermediates in the initial stages of epothilone biosynthesis. *Chem Biol* **11**, 327-335, doi:10.1016/j.chembiol.2004.02.021 (2004).

30 Jung, J., Bashiri, G., Johnston, J. M. & Baker, E. N. Mass spectral determination of phosphopantetheinylation specificity for carrier proteins in Mycobacterium tuberculosis. *FEBS Open Bio* **6**, 1220-1226, doi:10.1002/2211-5463.12140 (2016).

31 Miyanaga, A., Iwasawa, S., Shinohara, Y., Kudo, F. & Eguchi, T. Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proc Natl Acad Sci U S A* **113**, 1802-1807, doi:10.1073/pnas.1520042113 (2016).

32 Spizizen, J. Transformation of Biochemically Deficient Strains of Bacillus Subtilis by Deoxyribonucleate. *Proc Natl Acad Sci U S A* **44**, 1072-1078 (1958).

33 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97**, 6640-6645, doi:10.1073/pnas.120163297 (2000).

34 Yan, X., Yu, H. J., Hong, Q. & Li, S. P. Cre/lox system and PCR-based genome engineering in Bacillus subtilis. *Appl Environ Microbiol* **74**, 5556-5562, doi:10.1128/AEM.01156-08 (2008).

35 Cronan, J. E. & Thomas, J. Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. *Methods Enzymol* **459**, 395-433, doi:S0076-6879(09)04617-5 [pii]10.1016/S0076-6879(09)04617-5 (2009).

36 Layne, E. in *Methods in Enzymology* Vol. 3 447-454 (Academic Press, 1957).

37 Jackowski, C. O. R. Regulation of phospholipid synthesis in Escherichia coli. Composition of the acyl-acyl carrier protein pool in vivo. (1982).

38 Manandhar, M. & Cronan, J. E. Proofreading of noncognate acyl adenylates by an acyl-coenzyme a ligase. *Chem Biol* **20**, 1441-1446, doi:10.1016/j.chembiol.2013.10.010 (2013).

39 Christensen, Q. H., Martin, N., Mansilla, M. C., de Mendoza, D. & Cronan, J. E. A novel amidotransferase required for lipoic acid cofactor assembly in Bacillus subtilis. *Mol Microbiol* **80**, 350-363, doi:10.1111/j.1365-2958.2011.07598.x (2011).

40 Martin, N., Christensen, Q. H., Mansilla, M. C., Cronan, J. E. & de Mendoza, D. A novel two-gene requirement for the octanoyltransfer reaction of Bacillus subtilis lipoic acid biosynthesis. *Mol Microbiol* **80**, 335-349, doi:10.1111/j.1365-2958.2011.07597.x (2011).

41 Shivani, G., Huanan, J. I. N., Marna, Y.-N., Basil, J. N. & Iowa State University Research Foundation, I. Materials and methods for characterizing and using kasiii for production of bi-functional fatty acids. (2014).

42 Neidhardt, F. C. *et al. Escherichia coli and Salmonella typhimurium: cellular and molecular biology*. Vol. 2 (American Society for Microbiology Washington, DC, 1987).

43 Toomey, R. E. & Wakil, S. J. Studies on Mechanism of Fatty Acid Synthesis .15. Preparation and General Properties of Beta-Ketoacyl Acyl Carrier Protein Reductase from Escherichia Coli. *Biochim Biophys Acta* **116**, 189-& (1966).

44 Cronan, J. E. Biotin and lipoic acid: synthesis, attachment and regulation. *EcoSal Plus* **6** (2014).

45 Magnuson, K., Jackowski, S., Rock, C. O. & Cronan, J. E. Regulation of Fatty-Acid Biosynthesis in Escherichia-Coli. *Microbiol Rev* **57**, 522-542 (1993).

46 Persson, B. *et al.* The SDR (short-chain dehydrogenase/reductase and related enzymes) nomenclature initiative. *Chem-Biol Interact* **178**, 94-98 (2009).

47 Price, A. C., Zhang, Y.-M., Rock, C. O. & White, S. W. Structure of β-Ketoacyl-[acyl carrier protein] Reductase from Escherichia coli: Negative Cooperativity and Its Structural Basis†. *Biochemistry* **40**, 12772-12781 (2001).

48 Zhang, Y. & Cronan, J. E. Transcriptional analysis of essential genes of the Escherichia coli fatty acid biosynthesis gene cluster by functional replacement with the analogous Salmonella typhimurium gene cluster. *J Bacteriol* **180**, 3295-3303 (1998).

49 Lai, C.-Y. & Cronan, J. E. Isolation and characterization of β-ketoacyl-acyl carrier protein reductase (fabG) mutants of Escherichia coli and Salmonella enterica serovar Typhimurium. *J Bacteriol* **186**, 1869-1878 (2004).

50 Dill, K. A. *et al.* Principles of Protein-Folding - a Perspective from Simple Exact Models. *Protein Science* **4**, 561-602 (1995).

51 Matsui, P., DePaulo, J. & Buratowski, S. An interaction between the Tfb1 and SsI1 subunits of yeast TFIIH correlates with DNA repair activity. *Nucleic acids research* **23**, 767-772 (1995).

52 D'Argenio, D. A., Vetting, M. W., Ohlendorf, D. H. & Ornston, L. N. Substitution, insertion, deletion, suppression, and altered substrate specificity in functional protocatechuate 3,4-dioxygenases. *J Bacteriol* **181**, 6478-6487 (1999).

53 Beall, B. & Lutkenhaus, J. Sequence-Analysis, Transcriptional Organization, and Insertional Mutagenesis of the Enva Gene of Escherichia-Coli. *J Bacteriol* **169**, 5408-5415 (1987).

54 Bouquin, N., Tempete, M., Holland, I. B. & Seror, S. J. Resistance to Trifluoroperazine, a Calmodulin Inhibitor, Maps to the Fabd Locus in Escherichia-Coli. *Mol Gen Genet* **246**, 628-637 (1995).

55 Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**, 233-239, doi:10.1038/nbt.2508 (2013).

56 Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195-201, doi:10.1093/bioinformatics/bti770 (2006).

57 Guex, N., Peitsch, M. C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **30 Suppl 1**, S162-173, doi:10.1002/elps.200900140 (2009).

58 Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L. & Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res* **37**, D387-392, doi:10.1093/nar/gkn750 (2009).

59 Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* **42**, W252-258, doi:10.1093/nar/gku340 (2014).

60 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38 (1996).

61 LaRossa, R. A. & Van Dyk, T. K. Physiological roles of the DnaK and GroE stress proteins: catalysts of protein folding or macromolecular sponges? *Mol Microbiol* **5**, 529-534 (1991).

62 Van Dyk, T. K., Gatenby, A. A. & LaRossa, R. A. Demonstration by genetic suppression of interaction of GroE products with many proteins. *Nature* **342**, 451-453, doi:10.1038/342451a0 (1989).

63 Price, A. C., Zhang, Y. M., Rock, C. O. & White, S. W. Cofactor-induced conformational rearrangements establish a catalytically competent active site and a proton relay conduit in FabG. *Structure* **12**, 417-428 (2004).

64 Caroff, M., Tacken, A. & Szabo, L. Detergent-Accelerated Hydrolysis of Bacterial-Endotoxins and Determination of the Anomeric Configuration of the Glycosyl Phosphate Present in the Isolated Lipid-a Fragment of the Bordetella-Pertussis Endotoxin. *Carbohyd Res* **175**, 273-282 (1988).

65 Raetz, C., Purcell, S., Meyer, M., Qureshi, N. & Takayama, K. Isolation and characterization of eight lipid A precursors from a 3-deoxy-D-manno-octylosonic acid-deficient mutant of Salmonella typhimurium. *J Biol Chem* **260**, 16080-16088 (1985).

66 Sheldon, P. S., Kekwick, R. G. O., Smith, C. G., Sidebottom, C. & Slabas, A. R. 3-Oxoacyl-[Acp] Reductase from Oilseed Rape (Brassica-Napus). *Biochim Biophys Acta* **1120**, 151-159 (1992).

67 Dutta, D., Bhattacharyya, S., Roychowdhury, A., Biswas, R. & Das, A. K. Crystal structure of hexanoyl-CoA bound to beta-ketoacyl reductase FabG4 of Mycobacterium tuberculosis. *Biochem J* **450**, 127-139, doi:10.1042/BJ20121107 (2013).

68 Miller, D. J., Zhang, Y. M., Rock, C. O. & White, S. W. Structure of RhlG, an essential beta-ketoacyl reductase in the rhamnolipid biosynthetic pathway of Pseudomonas aeruginosa. *J Biol Chem* **281**, 18025-18032, doi:10.1074/jbc.M601687200 (2006).

69 Cigana, C. *et al.* Pseudomonas aeruginosa Exploits Lipid A and Muropeptides Modification as a Strategy to Lower Innate Immunity during Cystic Fibrosis Lung Infection. *PLoS ONE* **4**, e8439, doi:10.1371/journal.pone.0008439 (2009).

70 Dotson, G. D., Kaltashov, I. A., Cotter, R. J. & Raetz, C. R. Expression cloning of a Pseudomonas gene encoding a hydroxydecanoyl-acyl carrier protein-dependent UDP-GlcNAc acyltransferase. *J Bacteriol* **180**, 330-337 (1998).

71 Zeng, D. *et al.* Mutants resistant to LpxC inhibitors by rebalancing cellular homeostasis. *J Biol Chem* **288**, 5475-5486, doi:10.1074/jbc.M112.447607 (2013).

72 Caughlan, R. E. *et al.* Mechanisms decreasing in vitro susceptibility to the LpxC inhibitor CHIR-090 in the gram-negative pathogen Pseudomonas aeruginosa. *Antimicrob Agents Chemother* **56**, 17-27, doi:10.1128/AAC.05417-11 (2012).

73 Cronan, J. E. pBR322 vectors having tetracycline-dependent replication. *Plasmid* **84-85**, 20-26, doi:10.1016/j.plasmid.2016.02.004 (2016).

74 Gil, D. & Bouche, J. P. ColE1-type vectors with fully repressible replication. *Gene* **105**, 17-22 (1991).

75 Osawa, R. *et al.* Genotypic variations of Shiga toxin-converting phages from enterohaemorrhagic Escherichia coli O157: H7 isolates. *J Med Microbiol* **49**, 565-574, doi:10.1099/0022-1317-49-6-565 (2000).

76 Schicklmaier, P., Moser, E., Wieland, T., Rabsch, W. & Schmieger, H. A comparative study on the frequency of prophages among natural isolates of Salmonella and Escherichia coli with emphasis on generalized transducers. *Antonie Van Leeuwenhoek* **73**, 49-54 (1998).

77 Casjens, S. Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* **49**, 277-300 (2003).

78 Rotman, E., Amado, L. & Kuzminov, A. Unauthorized horizontal spread in the laboratory environment: the tactics of Lula, a temperate lambdoid bacteriophage of Escherichia coli. *PLoS One* **5**, e11106, doi:10.1371/journal.pone.0011106 (2010).

79 Rotman, E., Kouzminova, E., Plunkett, G., 3rd & Kuzminov, A. Genome of Enterobacteriophage Lula/phi80 and insights into its ability to spread in the laboratory environment. *J Bacteriol* **194**, 6802-6817, doi:10.1128/JB.01353-12 (2012).

80 Devoret, R., Pierre, M. & Moreau, P. Prophage ψ80 is induced in Escherichia coli K12 recA430. *Molecular and General Genetics MGG* **189**, 199-206 (1983).

81 Irbe, R. M., Morin, L. & Oishi, M. Prophage (phi 80) induction in Escherichia coli K-12 by specific deoxyoligonucleotides. *Proceedings of the National Academy of Sciences* **78**, 138-142 (1981).

82 Braun, V. FhuA (TonA), the career of a protein. *J Bacteriol* **191**, 3431-3436, doi:10.1128/JB.00106-09 (2009).

83 Frost, G. E. & Rosenberg, H. Relationship between the tonB locus and iron transport in Escherichia coli. *J Bacteriol* **124**, 704-712 (1975).

84 Killmann, H., Videnov, G., Jung, G., Schwarz, H. & Braun, V. Identification of receptor binding sites by competitive peptide mapping: phages T1, T5, and phi 80 and colicin M bind to the gating loop of FhuA. *J Bacteriol* **177**, 694-698 (1995).

85 Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2**, doi:10.1038/msb4100050 (2006).

86 Miller, J. (Cold Spring Harbor, NY, 1972).

87 Postle, K. in *Methods in enzymology* Vol. 422 245-269 (Elsevier, 2007).

88 Leong, J. M. *et al.* The phi 80 and P22 attachment sites. Primary structure and interaction with Escherichia coli integration host factor. *J Biol Chem* **260**, 4468-4477 (1985).

89 Franklin, N. C., Dove, W. F. & Yanofsky, C. The linear insertion of a prophage into the chromosome of E. coli shown by deletion mapping. *Biochemical and Biophysical Research Communications* **18**, 910-923 (1965).

90 Matsushiro, A. Specialized transduction of tryptophan markers in Escherichia coli K12 by bacteriophage∅ 80. *Virology* **19**, 475-482 (1963).

91 Kholodii, G. Y. & Mindlin, S. Integration of bacteriophages λ and ϕ80 in wild-type Escherichia coli at secondary attachment sites. *Molecular and General Genetics MGG* **197**, 104-108 (1984).

APPENDIX A

COPING WITH INADVERTENT LYSIS OF *ESCHERICHIA COLI* CULTURES: GENERATION OF STRAINS RESISTANT TO LYSOGENIZATION AND INFECTION BY THE STEALTHY LYSOGENIC BACTERIOPHAGE Φ80

Phages are the most abundant biological entities on Earth and understanding their pivotal role in a variety of biological and environmental processes has led to some of our most meaningful advances in molecular biology. Most of the known phages are virulent and invariably lyse their hosts in their normal life cycle. However, some phages, called lysogenic (temperate) phages, can remain as quiescent integrants (prophage) of the host replicon that are passively replicated by the host until conditions become favorable for lytic growth. These prophages are surprisingly common, with one study finding 51 different functional phages released from 27 different *Escherichia coli* strains [75] and another showing that 83 of 107 *E. coli* strains released at least one functional phage type [76]. Ultimately, these phage elements can account for up to 20% of the genomic DNA of their hosts [77].

Recent reports have described a lysogenic bacteriophage (Lula, later shown to be Φ80) that resides in many laboratory *Escherichia coli* strains [78,79]. Phage Φ80 infects other *E. coli* strains and spreads quickly and is generally not detected in aseptic laboratory conditions due to its strong propensity to lysogenize [78]. However, Φ80 lysogens are much less stable than those of phage λ, the paradigm lysogenic phage, [79-81]and spontaneously release large numbers of infectious phage particles. The presence of Φ80 phage particles in the medium of growing cultures is due to

induction by spontaneous DNA damage [78,79] whereas induction of phage λ lysogens requires much higher levels of DNA damage. The high titers of infectious Φ80 phage in the growth medium often accumulate without apparent lysis since most of the cells remain lysogens. The infectious virions are readily spread because they are stable in aerosols and to desiccation. Regardless, Φ80 lysogens have avoided detection, persisted for decades and are more prevalent than one might suspect. One reason, noted above, is the Φ80 preference for lysogeny upon initial infection. The second important factor is that Φ80 lytic development is very temperature dependent. Phage production is rapid at <30°C, slow at 37°C (the temperature at which *E. coli* strains are commonly grown), and is essentially blocked at 42°C. Indeed, lysis due to Φ80 is most commonly seen in cultures grown at 28-30°C, the permissive temperature of the temperature sensitive (Ts) strains used to study function of genes essential for growth such as those of DNA replication and membrane lipid synthesis. Moreover, growth of *E. coli* strains at temperatures <30°C is often used in high-level protein production. Indeed, the ability of Φ80 to flourish at low growth temperatures was key to its detection. We have observed Φ80 phage lysis of *E. coli* cultures in our laboratory and other laboratories report this extends to some strains from the *E. coli* Genetic Stock Center. We have sought to alleviate this problem by generating and characterizing a set of commonly used *E. coli* strains that cannot be infected. These, otherwise 'wild type', *E. coli* strains should allow effective control of the spread of Φ80.

We first considered adsorption of Φ80 to its *E. coli* host. Phage Φ80 has been shown to require the outer-membrane receptor FhuA, also known as TonA, for adsorption to the host. FhuA is coupled to the energized cytoplasmic membrane through TonB and both are required for Φ80 adsorption to be irreversible [82-84]. We constructed *tonA* and *tonB* strains in *E. coli* strain BW25113 by

replacing the genes with a kanamycin resistance cassette using the method of Datsenko and Wanner [33,85] and primer sets del tonA UP/del tonA DN and del *tonB* UP/del *tonB* DN, respectively. A Φ80 stock was prepared by growing a Φ80 lysogen (BW25113::Φ80) at 37°C overnight and then lysing the culture by adding a few drops of chloroform and removal of the cell debris by centrifugation.. The supernatant was titered by plating serial dilutions on wild type BW25113 and assay of plaque forming units. To test susceptibility to Φ80 infection, exponentially growing cultures were incubated with a Φ80 phage stock (titer 10^9 pfu/ml) and plated on LB agar plates and incubated at 30°C overnight. Both *fhuA* (*tonA*) and *tonB* strains were resistant to Φ80 infection while a control non-lysogenic BW25113 strain showed complete lysis (Fig. A.1). However, it should be noted that *fhuA* (*tonA*) and *tonB* mutant strains are deficient in the uptake of ferric enterochelin, ferric citrate and vitamin B-12[82,83] and thus cannot be used to study iron metabolism and related processes such as oxidative stress mediated by the Fenton reaction. In early work *tonB* mutations were used to make strains resistant to phage T1 (hence the *ton* designation) which uses the same receptor as Φ80 (many of the reported infections by phage T1 may well have been Φ80 infections given that Φ80 was a more mainstream phage [86]. However, a disadvantage (now passé) of *tonB* strains was sensitivity to the chromic acid then used to clean glassware [87]. For this reason, *tonB* mutations were replaced by the *fhuA* (*tonA*) mutations which are present in several commercially available cloning and expression strains.

The primary bacterial attachment site (*attB80*) of Φ80 was mapped [88] to be near the *trp* locus. Comparison of the core sequence of *attB80* to the published sequence of the *E. coli* K-12 strain MG1655, located it to the last 19 nucleotides of the gene *yciI* (Fig. A.2B), which is directly adjacent to *tonB* (albeit divergently transcribed). The yciI gene encodes a non-essential protein of

unknown function and seems independently transcribed. Considering the genomic context of the integration site within *yciI* and its immediate neighbor *tonB*, the Δ*yciI759::Kan* mutant (henceforth referred to as Δ*yciI Kc*) available from the Keio collection was tested for its ability to be lysogenized by Φ80. The lysogeny test is based on the fact that Φ80 lysogens are immune to superinfection by the same phage. An exponentially growing culture of BW25113 was inoculated with a Φ80 phage stock (multiplicity of infection <1) and allowed to grow for 15 min at 37°C before being plated on LB-agar plates incubated at the same temperature overnight. Surviving colonies were grown to OD_{600} of 0.2 at 37°C and cross-streaked against a high-titer Φ80 phage stock. The growth of Δ*yciI Kc* beyond the cross-streak indicated that the strain was a Φ80 lysogen which blocked infection (Fig. A.3B). The medium of the Δ*yciI Kc* was also found to contain very high phage titers ($>10^7$), comparable to a wild type lysogen (Fig. A.3A).

These results motivated us to construct wild type laboratory strain(s) that could not be efficiently lysogenized. Two different strains were constructed (Fig. A.2A). In one (called BW25113 Δ*attΦ80*) the chromosomal attachment site was deleted. The other strain (*attΦ8**) had a mutated attachment site core region sequence that was altered without changing the amino acids encoded in the carboxy terminus of the YciI protein (Fig. A.2C). Both mutant strains were constructed as described above using primer sets del att UP/del *yciI* DN and mut att UP/del *yciI* DN (Fig. A.2A). When tested for lytic infectivity, both strains showed remarkably lower phage production (at least 1000-fold lower). Also, when tested for lysogeny, neither strain could grow beyond the cross-streak, implying that they could not form Φ80 lysogens. This was verified by PCR amplification across the *attB* site using primers Lys check F/Lys check R. The presence of an amplified band indicated that the site was intact (Fig. A.3C) whereas integration of the large phage genome

(46,150 bp) would result in the lack of a band due to the inability of PCR to pr4ogress across the phage genome. We also constructed a complete deletion of *yciI* leaving only its stop codon (Δ*yciI*::Kan) and a deletion of the entire genomic region from the end of *yciI* to the end of *tonB* (Δ*yciI-tonB*::Kan ; Fig. A.2A). The kanamycin resistance cassette is flanked by a pair of FRT sites and can be easily removed by expressing the FLP recombinase encoded on plasmid pCP20 [33] but the cassette was left intact to facilitate movement into other strains by P1 transduction.

Note that the *yciI-tonB* deletion in the chromosome of strain SW07 and thus this strain lacks both the chromosomal attachment site (*attB80*) and the TonB Φ80 outer membrane receptor. Hence, this strain can neither be infected by Φ80 nor efficiently form Φ80 lysogens. The presence of both mutations may seem redundant. However, note that Φ80 genomes can enter strain SW07 by means that do not require the TonB receptor, such as by phage P1 transduction using a phage stock grown on a Φ80 lysogen. Phage P1 readily packages prophages λ and Φ80 together with the flanking bacterial chromosomal segments [38,89,90]. Upon introduction of the Φ80 genome by P1 transduction the phage can be inserted into the *E. coli* chromosome either by the standard lysogenic pathway or by homologous recombination using the flanking host sequences. It should be noted that like phage λ, Φ80 can insert into secondary attachment sites [91]. However, integration into secondary attachment sites occurs at a very low frequency (~10^{-6} of the normal frequency of lysogeny). Moreover, excision of Φ80 from the secondary sites is very inefficient (10^{-4} -10^{-5}) relative to excision from *attB80* [91]. The combined low frequencies of these events argue that the secondary attachment sites are not a complicating factor.

In conclusion, the *yciI-tonB::kan* deletion can be readily moved to other *E. coli* strains to avoid lysis by phage Φ80 and prevent propagation and spreading of this robust phage to the environment.

A.1 TABLE AND FIGURES

Bacterial Strains		
Strains	**Relevant Genotype**	**Source**
BW25113	*lacI rrnBT14 ΔlacZWJ16 hsdR514 ΔaraBADAH33 ΔrhaBADLD78 rph-1*	CGSC #7636
BW25113::Φ80	Φ80 lysogen	Personal gift of A. Kuzminov
JW1243	BW25113 Δ*yciI*759::Kan ; Also called Δ*yciI* Kc	Keio collection
SW03	BW25113 *ΔtonA::Kan*	This work
SW04	BW25113 *ΔtonB::Kan*	This work
SW05	BW25113 Δ*attB80*::Kan	This work
SW05	BW25113 mut *attB*80::Kan	This work
SW06	BW25113 Δ*yciI*::Kan	This work
SW07	BW25113 Δ*yciI-tonB*::Kan	This work
Plasmids		
Plasmids	**Relevant Characteristics**	**Source**
pKD46	Amp^R *ori* (Ts), Arabinose-inducible expression of λ-Red genes	Datsenko and Wanner, 2000
pKD4	Amp^R *pir* dependent provides FRT-KanR-FRT	Datsenko and Wanner, 2000
Oligonucleotides		
Oligonucleotides	**Sequence**	
del tonA UP	ATCATTCTCGTTTACGTTATCATTCACTTTACATCAGAGATATA CCAATGGTGTAGGCTGGAGCTGCTTCG	
del tonA DN	GCACGGAAATCCGTGCCCCAAAAGAGAAATTAGAAACGGAAG GTTGCGGTCATATGAATATCCTCCTTAGT	
del *tonB* UP	ATTTAAAATCGAGACCTGGTTTTTCTACTGAAATGATTATGACT TCAATG CATATGAATATCCTCCTTAGT	
del *tonB* DN	CTGTTGAGTAATAGTCAAAAGCCTCCGGTCGGAGGCTTTTGAC TTTCTGCGTGTAGGCTGGAGCTGCTTCG	
del *yciI* UP	CCCTTTCAAATAACGTACTTTACAACTTTCCCGAACAAGGAGTT GTGCCCGTGTAGGCTGGAGCTGCTTCG	

Table A.1 Bacterial Strains, Plasmids and Oligonucleotides used in this study

Table A.1 (cont.)

mut att UP	ATGCTGACCCGTACGTAGCGGCAGGAGTCTATGAACACGTTTC AGTGAAGCCTTTCAAAAAGGTCTTTTGAGTGTAGGCTGGAGCT GCTTCG
del *yciI* DN	AAACAAGAACACGGTTGCAAAAACCGTGCCCTTAAATATTGAA TCTCTATCATATGAATATCCTCCTTAGT
del att UP	ATGC TGACCCGTAC GTAGCGGCAG GAGTCTATGA ACACGTTTCAGTGTGAGTGTAGGCTGGAGCTGCTTCG
Lys check F	CTTTATGCATGGCCTTAACAAAGAC
Lys check R	TATTATCTGCTTGTGGTGGTGAATG

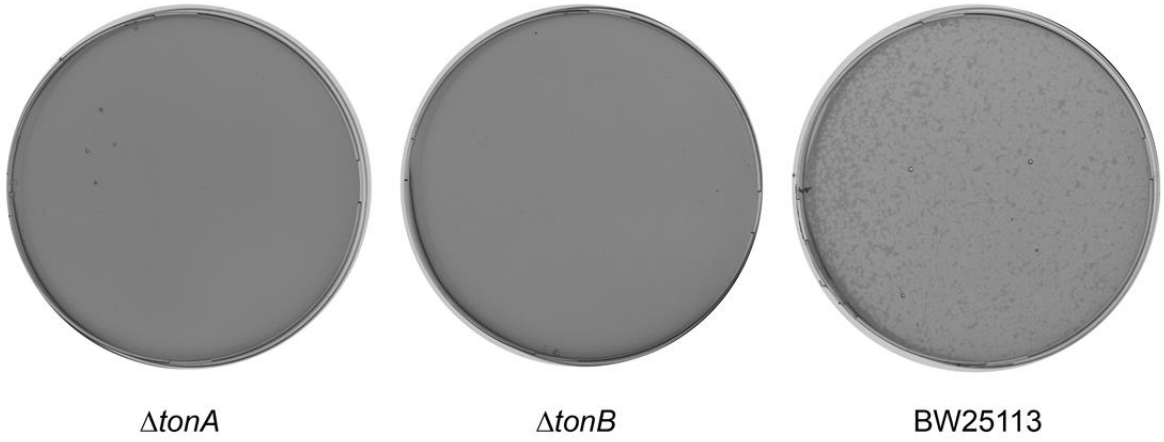

Figure A.1 Lysis by phage Φ80. Exponentially growing cultures of *E. coli* strains BW25113, BW25113 Δ*tonA* and BW25113 Δ*tonB* strains were infected with 10 µl of a phage Φ80 stock and plated on LB agar. BW25113 Δ*tonA* and BW25113 Δ*tonB* could not be infected by Φ80 as inferred from the lack of clearing. In contrast strain BW25113 plate showed almost confluent plaques (roughly 10^5 plaques)

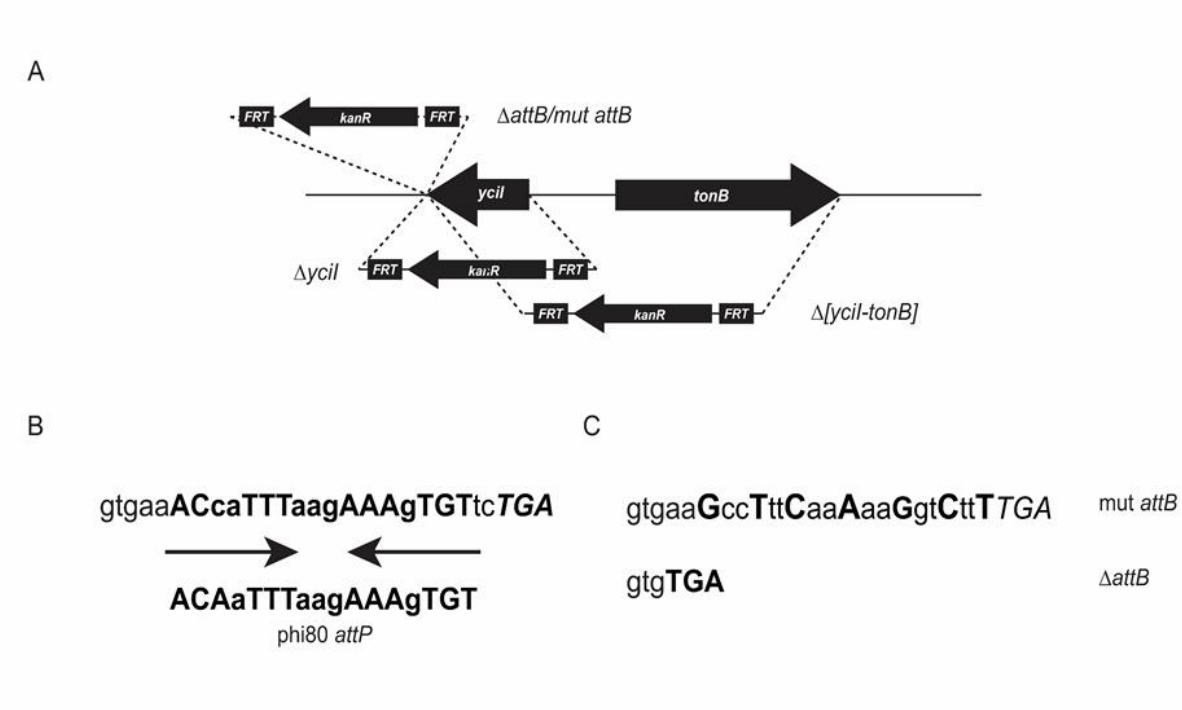

Figure A.2 Representations of A) Genetic manipulations performed in BW25113 B) The phage Φ80 bacterial attachment site (*attB*) and phage attachment site (*attP*) with homology indicated in capital letters and core region marked by arrows C) Nucleotide sequences of *attB* mutants. The nucleotide sequence of the core region has been sufficiently changed while preserving the amino acid sequence of *yciI* whereas in the Δ*yciI* strain the entire core region was deleted.

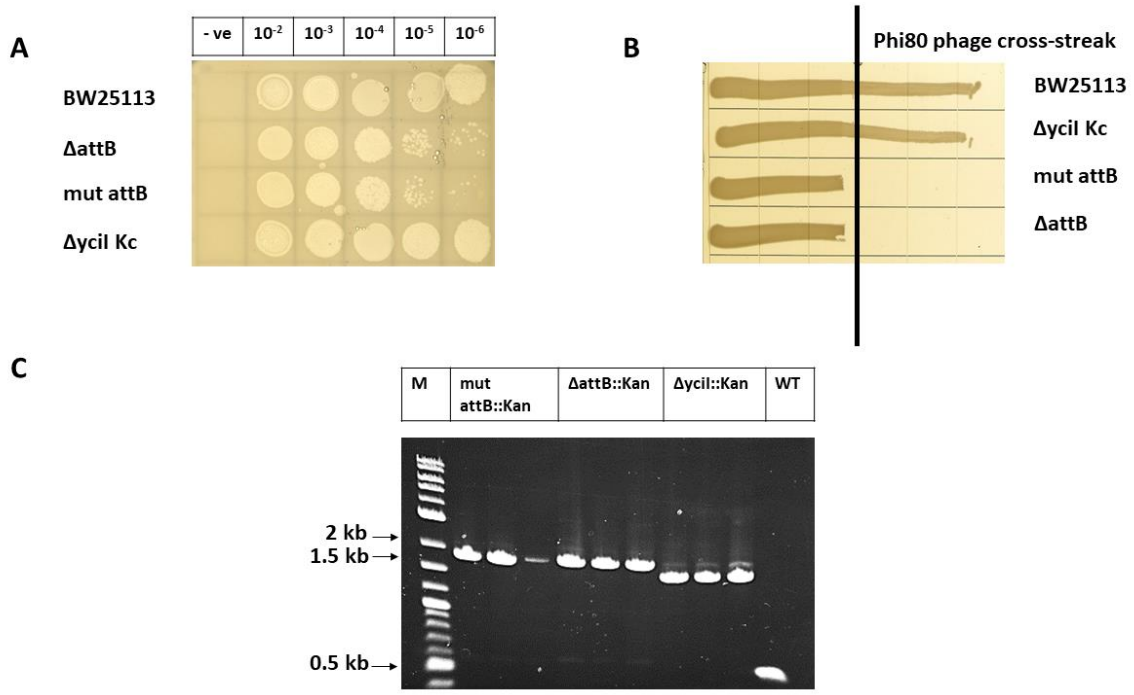

Figure A.3 Test of phage production and lysogeny. A) Phage titers were calculated by spotting 10 µl of ten-fold dilutions of the relevant phage stocks on top agar containing exponentially growing cells of susceptible BW25113 at 30˚C. Clear zones denote confluent lysis whereas distinct plaques formed on the last two dilutions of the *attB*80 mutant strains. B) Lysogenic strains were tested by streaking a culture across a thick streak of a phage stock. Lysogens survive due to their inability to be superinfected whereas non-lysogens are killed by phage infection. C) PCR amplification across the bacterial attachment site to verify lack of phage lysogeny. The larger sizes of the bands in the mutant strains relative to the wild type band is due to insertion of the kanamycin resistance determinant.